Received: 16 May 2019 | Revised: 7 January 2020 | Accepted: 30 January 2020 | Published on: 18 February 2020

DOI: 10.1002/nav.21892

RESEARCH ARTICLE

WILEY

A sequential partial information bomber-defender shooting problem

Krishna Kalyanam[1] | **David Casbeer**[2] | **Meir Pachter**[3]

[1]System Sciences Lab, PARC, Palo Alto, California
[2]Air Force Research Laboratory, Wright-Patterson AFB, Ohio
[3]Electrical & Computer Engineering Department, Air Force Institute of Technology, Wright-Patterson AFB, Ohio

Correspondence
Krishna Kalyanam, System Sciences Lab, PARC, Palo Alto, CA.
Email: krishnak@ucla.edu

Funding information
Air Force Research Laboratory, F48650-16-C-2642, Subcontract: 162642-19-18-C1.

Abstract

A bomber carrying homogenous weapons sequentially engages ground targets capable of retaliation. Upon reaching a target, the bomber may fire a weapon at it. If the target survives the direct fire, it can either return fire or choose to hold fire (play dead). If the former occurs, the bomber is immediately made aware that the target is alive. If no return fire is seen, the true status of the target is unknown to the bomber. After the current engagement, the bomber, if still alive, can either re-engage the same target or move on to the next target in the sequence. The bomber seeks to maximize the expected cumulative damage it can inflict on the targets. We solve the perfect and partial information problems, where a target always fires back and sometimes fires back respectively using stochastic dynamic programming. The perfect information scenario yields an appealing threshold based bombing policy. Indeed, the marginal future reward is the threshold at which the control policy switches and furthermore, the threshold is monotonic decreasing with the number of weapons left with the bomber and monotonic nondecreasing with the number of targets left in the mission. For the partial information scenario, we show via a counterexample that the marginal future reward is not the threshold at which the control switches. In light of the negative result, we provide an appealing threshold based heuristic instead. Finally, we address the partial information game, where the target can choose to fire back and establish the Nash equilibrium strategies for a representative two target scenario.

KEYWORDS

attacker-defender game, partial information, sequential decision making, shoot-look-shoot

1 | INTRODUCTION

The operational scenario is the following. A bomber with M identical weapons travels along a designated route/path and sequentially encounters N enemy targets on the ground. Upon reaching a target, the bomber may choose to release a weapon. A weapon dropped on a target will destroy it with probability p, where $0 < p < 1$. Successful elimination of the target yields a known positive reward to the bomber/decision maker (DM). However, we assume that the target is equipped with a surface to air missile launcher and so, is capable of firing back at the bomber. We assume the following sequence of events. The bomber acts first and fires at the target. If a weapon dropped on a target is unsuccessful, the target (which is still alive) can fire back at the bomber. We assume the probability that the bomber is not destroyed by a round of return fire is given by $s < 1$. After each engagement, if still alive, the bomber can either re-engage the current target or move on to the next target in the sequence. We are interested in the optimal weapon allocation policy that results in maximal total expected reward for the bomber. We emphasize here that the actual reward accrued (realization of a random variable) is not known to the bomber and will perhaps be collected by some form of ground intelligence, for example, by checking post-mission what targets were destroyed. The only exception is the perfect

This work was presented in part at the Second IMA & OR Society Conference on Mathematics of OR, Birmingham, UK, April 2019.

information scenario, where the target by virtue of firing back (or not) reveals the reward yielded to the bomber. The bomber's decision is clearly a function of the information regarding the target status (alive or dead). We consider three different but related information models in this article:

1. *Perfect information*: If attacked and not destroyed, the target always fires back. Here, the DM is immediately made aware of the true status of the target upon observing the presence or lack of return fire.
2. *Partial information*: If attacked and not destroyed, the target fires back with probability $f < 1$ known to the DM. If return fire is seen, the DM knows that the target is still alive. However, if there is no return fire, the DM cannot distinguish between the two possible states (alive/dead) of the target.
3. *Partial information game*: If attacked and not destroyed, the target can either choose to fire back or play dead (hold fire). Again, if there is no return fire, the DM cannot determine the true state of the target.

1.1 | Prior work

The framework we consider is common to military weapon target engagements (Aviv & Kress, 1997; Kisi, 1976; Mastran & Thomas, 1973), shooting problems (Glazebrook & Washburn, 2004; Sato, 1997a), cybersecurity (Gao, Zhong, & Mei, 2013; Hu, Xu, Xu, & Zhao, 2017), attacker-defender games (Levitin & Hausken, 2012), and more broadly to stochastic sequential resource allocation (Sato, 1996, 1997b). In sequential assignment problems of the kind considered herein, the decision rule usually takes the form, *attack the target iff its value is no less than a certain threshold c*, where the optimal c is to be determined. Moreover, one would expect the optimal threshold to be monotonic decreasing and nondecreasing in the number of remaining weapons and targets respectively. In other words, if the DM has more weapons in hand or less targets to engage, it is more likely to bomb the current target. If the probability of kill, $p = 1$, there is no need for repeated allocation to the same target and the resulting problem bears similarity to revenue management (RM)—wherein, the threshold monotonicity property is known to hold, for example, see (Aydin, Akçay, & Karaesmen, 2009; van Ryzin & Talluri, 2005), chap. 6). Monotonicity properties A, B, and C for the related *bomber problem*, where a bomber has to survive sequential engagements with enemy aircraft by firing a volley (one or more) of weapons are discussed in Weber (2013). Note that our setup is different from the bomber problem in two respects. In the bomber problem, (1) a volley of weapons (≥ 1) is allowed at each decision stage and (2) the enemy aircraft shoots at the bomber regardless of whether or not it was shot at first. The property B states that the optimal number of weapons (salvo size) assigned to an enemy aircraft is nondecreasing in the number of weapons left with the bomber. This property, long thought to be true, remained an open problem for 43 years. Recently, it was shown to be false via counterexamples (Kamihigashi, 2017).

To clearly distinguish this paper from our earlier work, which only considered passive targets with no retaliatory action, we note that:

1. In Kalyanam, Rathinam, Casbeer, and Pachter (2016), a perfect information scenario is considered, where the DM is informed if the target was destroyed or not and the threshold monotonicity result has been established for this case.
2. In Kalyanam, Casbeer, and Pachter (2017), the rewards are considered to be random (as in RM) but known at decision time and the target status is perfectly known to the DM; the threshold monotonicity result has been established for this case as well.
3. In Kalyanam, Casbeer, and Pachter (2018), we extend the monotonicity result to the case of error-prone battle damage assessment, where a live target is sometimes reported to be dead and vice-versa, that is, the DM has access to a classifier with known Type I and II error rates.

In military parlance, our model embraces a shoot-look-shoot (SLS) approach, in that homogenous resources are expended by the DM one at a time with observations made in between that assess the outcome of the previous allocation. The analysis of SLS strategies under partial feedback information on the allocation outcome was first reported in Aviv and Kress (1997). A survey of the state of the art in SLS methods and the optimal use of information are provided in Glazebrook and Washburn (2004). If additional complicating factors such as a search cost for finding a target or a scenario wherein resources can be replenished are considered, the monotonicity property breaks down, for example, see Sato (1996, 1997b). A related perfect information sequential allocation game, where an attacker and a defender choose to expend a single resource from a finite inventory is presented in Sakaguchi (1977).

A more elaborate partially observable Markov model where a single red engages multiple blue entities (possibly of different types) is provided in Glazebrook, Mitchell, Gaver, and Jacobs (2004). The authors therein analyze the optimal shooting strategy via generalized bandits. In contrast, our parsimonious model requires only two readily available parameters, the probabilities of kill for the direct and return fire. This yields a fairly elegant solution derived entirely from first principles.

In this article, we present a low resolution model of the SLS scenario where the target can defend itself by returning fire at the DM. Our emphasis is on providing an analytical relationship under which it is optimal for the target to play dead. We also derive critical threshold monotonicity properties for the optimal policy under perfect information and show that a similar result does not hold for the partial information

scenario. Motivated by the perfect information policy, we also provide an intuitive threshold based heuristic policy for the partial information case supported by numerical simulations.

2 | PERFECT INFORMATION SCENARIO

We assume that the bomber is equipped with M identical weapons and sequentially visits N targets on the ground. Under perfect information, we stipulate that if the bomber chooses to deploy a weapon and if the (current) target is not destroyed, the target immediately fires back at the bomber. We emphasize that the bomber *looks* to see if the target fires back or not before proceeding with the next course of action (SLS approach). Let $V(j, k)$ indicate the optimal cumulative reward (value function) that can be achieved when the DM (still alive) arrives at the jth live target with k weapons in hand. It follows that the value function must satisfy the Bellman recursion:

$$\begin{aligned} V(j,k+1) = \max_{u=0,1}\{&[p(r_j + V(j+1,k)) + qsV(j,k)]u \\ &+ (1-u)V(j+1,k+1)\}, \\ &j = 1, \ldots, N-1, \quad k = 0, \ldots, M-1, \end{aligned} \tag{1}$$

where $q = 1-p$. The control action, u indicates whether the DM should deploy a weapon at the current target ($u = 1$) or move on to the next target ($u = 0$). The decision $u = 1$ results in the jth target being destroyed with probability p yielding an immediate reward of r_j and a future expected payoff of $V(j+1, k)$. In other words, having successfully destroyed target j, the DM moves on to target $j+1$ with k weapons in hand. On the other hand, if it is not destroyed, the target fires back at the DM. In this case, the DM survives with probability $s < 1$ and so, the corresponding expected future payoff is $sV(j, k)$. If $u = 0$ is chosen, the DM simply moves on to the next target in the sequence and the corresponding expected future payoff is $V(j+1, k+1)$.

In lieu of (1), the optimal firing policy is given by:

$$\begin{aligned} \mu(j,k+1) = \arg\max_{u=0,1}\{&[p(r_j + V(j+1,k)) + qsV(j,k)]u \\ &+ (1-u)V(j+1,k+1)\}, \\ &j = 1, \ldots, N-1, \quad k = 0, \ldots, M-1. \end{aligned} \tag{2}$$

If the DM runs out of ammunition, there is no more reward to be gained that is, the boundary condition:

$$V(j,0) = 0, \quad \forall j. \tag{(3)}$$

Furthermore, if the DM is alive at the last target (known to be alive) and has $k > 0$ weapons at hand, it is always optimal to fire a weapon and so, the Bellman recursion (1) collapses to:

$$\begin{aligned} &V(N,k) = pr_N + qsV(N,k-1), \\ \Rightarrow\ &V(N,k) = pr_N \sum_{i=0}^{k-1} (qs)^i \\ &\qquad = \frac{pr_N}{1-qs}[1-(qs)^k], \quad k = 1, \ldots, M. \end{aligned} \tag{4}$$

It is clear that the optimal value function can be easily computed using backward dynamic programming. The problem is tractable so long as M, N are small as can be expected in a realistic military engagement. In this article, we are interested more in the structure of the optimal solution. In particular, we are interested in showing that a (monotonic) thresholding policy is optimal.

2.1 | Threshold firing policy

If the bomber chooses to move on to the next target, the expected reward is $V(j+1, k+1)$. On the other hand, if it chooses to deploy a weapon at the current target j, it must either destroy the target or evade the return fire to remain alive. So, the bomber will survive the current engagement with probability $p + qs$ and thereafter, it will receive the reward $V(j+1, k)$ from downstream targets. Hence, the marginal future reward is given by,

$$\begin{aligned} \Delta_{j+1}(k) \coloneqq V(j+1,k+1) - (p+qs)V(j+1,k), \\ j = 1, \ldots, (N-1). \end{aligned} \tag{5}$$

Suppose $\Delta_{j+1}(k)$ is a monotonically decreasing function of k for all $j = 1, \ldots, (N-1)$. Let $\kappa(j)$ be the smallest nonnegative integer k such that $pr_j \geq \Delta_{j+1}(k)$ for all $j = 1, \ldots, (N-1)$. The following result shows that, under this assumption, a thresholding policy is optimal for the DM.

Proposition 1 *If $\Delta_{j+1}(k)$ is a monotonically decreasing function of k, we have for all $j < N$,*

$$\mu(j,k) = \begin{cases} 0 & k \leq \kappa(j), \\ 1 & \textit{otherwise}. \end{cases}$$

Proof From the definition of $\Delta_{j+1}(k)$ and $\kappa(j)$, we have:

$$\begin{aligned} pr_j < \Delta_{j+1}(k), \ \ k < \kappa(j) \quad \text{and} \quad pr_j \geq \Delta_{j+1}(k), \\ k \geq \kappa(j). \end{aligned} \tag{6}$$

From the Bellman recursion (1), we have: $V(j, k) \geq V(j+1, k)$. Therefore, it follows that:

$$\begin{aligned} p(r_j + V(j+1,k)) + qsV(j,k) \geq pr_j \\ + (p+qs)V(j+1,k). \end{aligned} \tag{7}$$

Combining (6) and (7), we can write:

$$\begin{aligned} p(r_j + V(j+1,k)) + qsV(j,k) \geq V(j+1,k+1), \\ \forall k \geq \kappa(j). \end{aligned} \tag{8}$$

In light of (8), the Bellman recursion (1) yields:

$$\begin{aligned} V(j,k+1) = p(r_j + V(j+1,k)) + qsV(j,k), \\ \forall k \geq \kappa(j), \end{aligned} \tag{9}$$

$$\Rightarrow \mu(j,k+1) = 1, \quad \forall k \geq \kappa(j). \tag{10}$$

We shall show that $\mu(j, k+1) = 0,\ \forall\, k < \kappa(j)$, by induction on k. Recall that:

$$pr_j + (p+qs)V(j+1,k) < V(j+1,k+1), \quad \forall k < \kappa(j). \tag{11}$$

If $\kappa(j) = 0$, there is nothing left to prove. So, suppose $\kappa(j) > 0$. From the Bellman recursion (1), we have:

$$\begin{aligned} V(j,1) &= \max_{u=0,1}\{pr_j u + (1-u)V(j+1,1)\} \\ &= V(j+1,1), \end{aligned} \tag{12}$$

where (12) follows from (11) applied to the case $k = 0$. So, we have: $\mu(j, 1) = 0$.

Suppose $V(j, h) = V(j+1, h)$ for some $h < \kappa(j)$. The Bellman recursion (1) yields:

$$\begin{aligned} V(j,h+1) &= \max_{u=0,1}\{[p(r_j + V(j+1,h)) + qsV(j,h)]u \\ &\quad + (1-u)V(j+1,h+1)\} \\ &= \max_{u=0,1}\{[pr_j + (p+qs)V(j+1,h)]u \\ &\quad + (1-u)V(j+1,h+1)\} \\ &= V(j+1,h+1). \end{aligned}$$

$$\Rightarrow \mu(j,h+1) = 0. \tag{13}$$

where (13) follows from applying (11) to the case $k = h$. Moreover, from (9) and (13), we can write:

$$\begin{aligned} V(j+1,\kappa(j)+1) &= p[r_j + V(j+1,\kappa(j))] + qsV(j,\kappa(j)) \\ &= p[r_j + V(j+1,\kappa(j))] + qsV(j+1,\kappa(j)) \\ &= pr_j + (p+qs)V(j+1,\kappa(j)). \end{aligned} \tag{14}$$

■

Combining (9), (13), and (14), we have:

$$V(j,k+1) = \begin{cases} V(j+1,k+1), & k < \kappa(j) \\ pr_j + (p+qs)V(j+1,k), & k = \kappa(j) \\ p(r_j + V(j+1,k)) + qsV(j,k), & k > \kappa(j). \end{cases} \tag{15}$$

Therefore, the corresponding optimal policy satisfies:

$$\mu(j,k+1) = \begin{cases} 0, & k < \kappa(j) \\ 1, & k \geq \kappa(j). \end{cases} \tag{16}$$

The above result tells us that 1 out of $(k+1)$ weapons is dropped on the current target j iff the immediate expected reward, pr_j is greater than or equal to the marginal future reward, $\Delta_{j+1}(k)$. So, the marginal future reward is the threshold at which the bomber's control switches. In line with previous results on the related discrete bomber problem, one would expect the optimal threshold to be monotonic decreasing in k and monotonic nonincreasing in j. These intuitive monotonicity properties are indeed true under our model as shown below.

Theorem 1 *For $j = 1, \ldots, N$, $\Delta_j(k)$ is a monotonic decreasing function of k.*

Proof We prove the result by backward induction on j. From (4) and (5), we have:

$$\begin{aligned} \Delta_N(k) &= V(N,k+1) - (p+qs)V(N,k) \\ &= \frac{pr_N}{1-qs}[1 - (qs)^{k+1} - (p+qs)(1-(qs)^k)] \\ &= \frac{pr_N}{1-qs}[q(1-s) + p(qs)^k]. \end{aligned} \tag{17}$$

$\Delta_N(k)$ is clearly a decreasing function of k given that $0 < q,\ s < 1$. Let us suppose that $\Delta_{j+1}(k)$ is a decreasing function of k for some $j < N-1$. As before, let $\kappa(j)$ be the smallest $k = 0, 1, \cdots$ such that $pr_j \geq \Delta_{j+1}(k)$. By definition, $\Delta_j(k) = V(j, k+1) - (p+qs)V(j, k)$ and so, we have from (15):

$$\Delta_j(k) = \Delta_{j+1}(k), \quad k < \kappa(j), \tag{18}$$

$$\Delta_j(\kappa(j)) = pr_j. \tag{19}$$

For $\ell = k - \kappa(j) > 0$, we have by repeated application of (15):

$$\begin{aligned} V(j,k+1) &= pr_j \sum_{i=0}^{\ell} (qs)^i + (qs)^{\ell}(p+qs)V(j+1,\kappa(j)) \\ &\quad + p\sum_{i=0}^{\ell-1} (qs)^i V(j+1,k-i). \end{aligned} \tag{20}$$

So, we can write:

$$\begin{aligned} \Delta_j(k) &= V(j,k+1) - (p+qs)V(j,k) \\ &= p\sum_{i=0}^{\ell-1} (qs)^i \Delta_{j+1}(k-i-1) \\ &\quad + pr_j\left[(qs)^{\ell} + q(1-s)\sum_{i=0}^{\ell-1}(qs)^i\right],\ k > \kappa(j). \end{aligned} \tag{21}$$

We proceed to show that $\Delta_j(k)$ as prescribed by (18), (19), and (21) is a decreasing function of k. By our induction argument, $\Delta_j(k) = \Delta_{j+1}(k)$ decreases as k goes from 0 to $\kappa(j) - 1$. From the definition of $\kappa(j)$, we have: $pr_j < \Delta_{j+1}(\kappa(j) - 1)$ and so,

$$\Delta_j(\kappa(j)) = pr_j < \Delta_{j+1}(\kappa(j)-1) = \Delta_j(\kappa(j)-1). \tag{22}$$

From (21), we have:

$$\begin{aligned} \Delta_j(\kappa(j)+1) &= p\Delta_{j+1}(\kappa(j)) + pr_j[qs + q(1-s)] \\ &= pr_j + p[\Delta_{j+1}(\kappa(j)) - pr_j] \end{aligned} \tag{23}$$

$$\leq \Delta_j(\kappa(j)), \tag{24}$$

since $\Delta_{j+1}(\kappa(j)) \leq pr_j$ as per the definition of $\kappa(j)$. For $k > \kappa(j)$ we again employ (21) and after

some algebraic manipulations get,

$$\Delta_j(k+1) - \Delta_j(k) = p(qs)^{\ell}[\Delta_{j+1}(\kappa(j)) - pr_j] + p\sum_{i=0}^{\ell-1}(qs)^i \times [\Delta_{j+1}(k-i) - \Delta_{j+1}(k-i-1)] < 0, \quad (25)$$

since $\Delta_{j+1}(k-i) < \Delta_{j+1}(k-i-1)$ per the induction argument and $\Delta_{j+1}(\kappa(j)) \le pr_j$. Hence, $\Delta_j(k)$ is a decreasing function of k. ∎

Remark 1 We note from (23) that strict monotonicity is guaranteed for $\Delta_j(k)$ as a function of k so long as $pr_j > \Delta_{j+1}(\kappa(j))$.

Next, we show that the threshold function is monotonic nondecreasing with the number of targets left in the mission.

Theorem 2 *For any k, $\Delta_j(k)$ is a monotonic nonincreasing function of j.*

Proof From (18) in Theorem 1, we have:

$$\Delta_j(k) = \Delta_{j+1}(k), \quad k = 0, \ldots, \kappa(j) - 1. \quad (26)$$

From (22) in Theorem 1 and the definition of $\kappa(j)$, we have:

$$\Delta_j(\kappa(j)) = pr_j \ge \Delta_{j+1}(\kappa(j)). \quad (27)$$

Recall that for $k > \kappa(j)$, $\Delta_j(k)$ is given by (21). From Theorem 1, $\Delta_{j+1}(k-i-1) > \Delta_{j+1}(k)$, $i \ge 0$ and so we have:

$$\begin{aligned}\Delta_j(k) &> \Delta_{j+1}(k)p\sum_{i=0}^{\ell-1}(qs)^i + pr_j\left[(qs)^{\ell} + q(1-s)\sum_{i=0}^{\ell-1}(qs)^i\right] \\ &= \Delta_{j+1}(k)[1-q(1-s+s)]\frac{1-(qs)^{\ell}}{1-qs} + pr_j\left[(qs)^{\ell} + q(1-s)\frac{1-(qs)^{\ell}}{1-qs}\right] \\ &= \Delta_{j+1}(k) + (pr_j - \Delta_{j+1}(k)) \times \left[(qs)^{\ell} + q(1-s)\frac{1-(qs)^{\ell}}{1-qs}\right] \quad (28)\end{aligned}$$

$$> \Delta_{j+1}(k), \quad \ell = k - \kappa(j) > 0. \quad (29)$$

In the above, (29) follows from the definition of $\kappa(j)$, that is, $pr_j > \Delta_{j+1}(k)$, $k > \kappa(j)$. From (26), (27), and (29), the result follows. ∎

Given the monotonicity result in Theorem 1, we can compute the optimal threshold at which the control switches via a direct backward linear recursion without resorting to the computationally prohibitive nonlinear Bellman recursion (1). Indeed, we can combine (18), (19), and (21) to get the backward recursion: $\forall j < N$,

$$\Delta_j(k) = \begin{cases} \Delta_{j+1}(k), & k < \kappa(j), \\ pr_j, & k = \kappa(j), \\ p\sum_{i=0}^{\ell-1}(qs)^i\Delta_{j+1}(k-i-1) + pr_j \times \left[(qs)^{\ell} + q(1-s)\sum_{i=0}^{\ell-1}(qs)^i\right]. \end{cases} \quad (30)$$

Recall that $\kappa(j)$ is the smallest $k = 0, 1, \ldots$ such that $pr_j \ge \Delta_{j+1}(k)$ for all $j < N$ and the boundary condition:

$$\Delta_N(k) = \frac{pr_N}{1-qs}[q(1-s) + p(qs)^k], \quad k \ge 0. \quad (31)$$

The backward recursion (30) yields the marginal reward values Δ_j and the corresponding threshold values $\kappa(j)$ for $j = 1, \ldots, (N-1)$. The optimal firing policy for all $j < N$ is prescribed by:

$$\mu(j,k) = \begin{cases} 0, & k \le \kappa(j), \\ 1, & \text{otherwise.} \end{cases} \quad (32)$$

At the last target, it is always optimal to fire a weapon and so, $\mu(N, k) = 0, \forall k > 0$. To visualize the central result, we have computed and plotted the marginal future reward function $\Delta_j(k)$ (top plot) and $\kappa(j)$ values (bottom plot) at which the control switches for six different sets of problem parameters in Figure 1. For illustration, we picked $N = 5$ targets. The other randomly chosen problem parameters, p, s, and r_j for different j are shown in the plots. The key takeaway is the dual monotonicity property of the marginal reward function. In the next section, we investigate if a similar threshold monotonicity property extends to the partial information case, when the target action is random.

3 | PARTIAL INFORMATION SCENARIO

In this section, we introduce an additional complexity in that the target, if attacked and not destroyed, fires back with probability $f \le 1$ known to the bomber. Suppose the DM is alive and initially facing a target known to be alive. Further suppose that subsequently $m \ge 1$ contiguous shots are fired at the target with no return fire seen. Let $\mathcal{P}_m$ indicate the corresponding a posteriori probability that the target is still alive after the stated sequence of events. We apply the conditional expectation theorem as follows. Let the events:

A: The target is alive after m rounds of contiguous firing and
B: No return fire was seen immediately after the mth round of firing.

It follows from the definition earlier that,

$$\mathcal{P}_m = Prob(\mathbf{A}|\mathbf{B}) = \frac{Prob(\mathbf{B}|\mathbf{A})Prob(\mathbf{A})}{Prob(\mathbf{B})}. \quad (33)$$

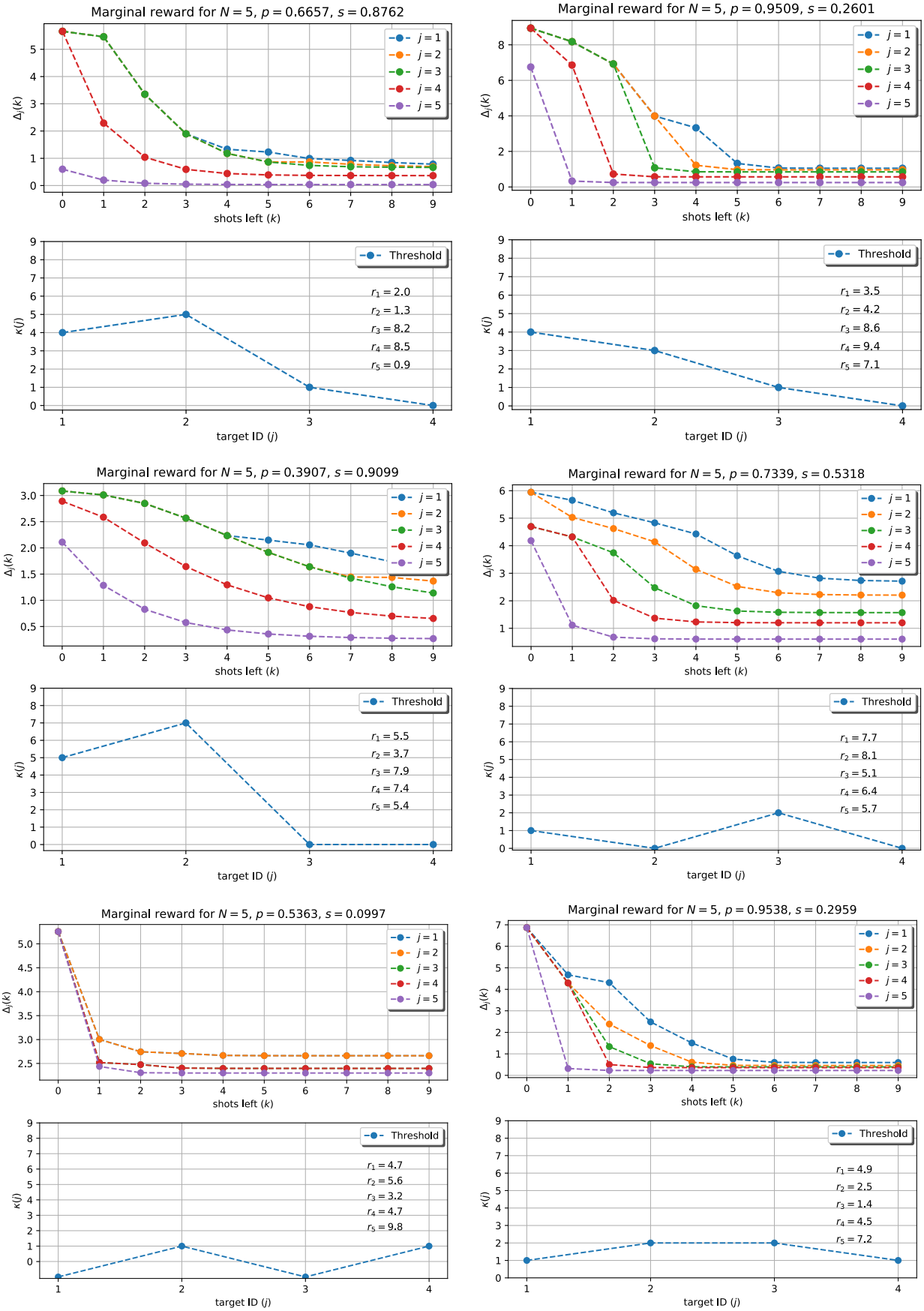

FIGURE 1 Monotonic marginal reward function and optimal threshold values for different problem parameters [Colour figure can be viewed at wileyonlinelibrary.com]

We note that $Prob(\mathbf{B}|\mathbf{A}) = 1-f$ is the probability that a live target holds fire. For the target to be alive after m rounds, it must have been alive after $m-1$ rounds and it must be missed by the mth shot. So, $Prob(\mathbf{A}) = \mathcal{P}_{m-1}q$. In addition, if the target must fire back, the corresponding probability is given by $\mathcal{P}_{m-1}qf$. It follows that the complementary event $\mathbf{B}$ satisfies: $Prob(\mathbf{B}) = 1 - \mathcal{P}_{m-1}qf$. Thus, the one step Bayes' update formula for the a posteriori probability is given by:

$$\mathcal{P}_m = \frac{\mathcal{P}_{m-1}q(1-f)}{1-\mathcal{P}_{m-1}qf}. \tag{34}$$

For $m=0$, we stipulate that $\mathcal{P}_0 = 1$. Note that this reflects the situation when the DM arrives for the first time at a live target or immediately after the DM survives a round of return fire from the target. In either case, the DM knows that the target is alive prior to any further engagement.

3.1 | Partial information sequential engagement

Suppose the DM is alive and initially facing (live) target j with k weapons in hand. Further suppose that $m \geq 0$ contiguous shots have been fired at j with no return fire seen thus far. Let $V(j,\, k \,|\, m)$ indicate the optimal expected cumulative reward that is, value function, that can be achieved thereafter. As before, let the control action, $u=1$ indicate that a shot is fired at the current target j and $u=0$ indicate that the DM moves on to the next target in the sequence. If the DM chooses $u=1$, it will observe the event $y=0,\,1$ indicating the absence or presence of return fire. In accordance with the SLS framework, we assume that the DM waits (*looks*) after firing a shot to see if there is return fire. Note that the target not firing back could indicate one of two things: (1) the target was destroyed or (2) the target was not destroyed *and* it did not fire back. These two states are indistinguishable to the DM. Indeed, we have the probability:

$$Prob(y=1) = \mathcal{P}_m qf, \tag{35}$$

where, as before, $q = 1-p$. In other words, for the target to fire back, it must be alive and choose to fire back. It follows that the value function must satisfy the Bellman recursion:

$$\begin{aligned} V(j,k+1 \mid m) = \max_{u=0,1}\{&[p\mathcal{P}_m r_j + Prob(y=1)sV(j,k \mid 0) \\ &+Prob(y=0)V(j,k \mid m+1)]u \\ &+(1-u)V(j+1,k+1 \mid 0)\}, \\ k=0,\ldots,(M-1), \quad &j=1,\ldots,(N-1). \end{aligned} \tag{36}$$

In (36), $u=1$ yields an expected immediate reward of $p\mathcal{P}_m r_j$ and the expected future payoff for the two outcomes $y=0,\,1$ are given by: $V(j,\, k \,|\, m+1)$ and $sV(j,\, k \,|\, 0)$ respectively. Indeed, if $u=1$ and $y=1$, the bomber must survive the return fire to yield any future reward, hence the factor s in front of $V(j,\, k \,|\, 0)$. If $u=1$ and $y=0$, an additional (contiguous) shot has been fired at j with no return fire and so, the corresponding expected future payoff is $V(j,\, k \,|\, m+1)$. As before, $u=0$ results in the DM moving on to the next target with the corresponding expected future payoff $V(j+1,\, k+1 \,|\, 0)$. Substituting for $Prob(y)$ from (35), we have:

$$\begin{aligned} V(j,k+1 \mid m) = \max_{u=0,1}\{&[p\mathcal{P}_m r_j + \mathcal{P}_m qfsV(j,k \mid 0) \\ &+[1-\mathcal{P}_m qf]V(j,k \mid m+1)]u \\ &+(1-u)V(j+1,k+1 \mid 0)\}, \\ k=0,\ldots,(M-1), \quad &j=1,\ldots,(N-1). \end{aligned} \tag{37}$$

The optimal firing policy is therefore given by:

$$\begin{aligned} \mu(j,k+1 \mid m) = \arg\max_{u=0,1}\,&[p\mathcal{P}_m r_j + \mathcal{P}_m qfsV(j,k \mid 0) \\ &+[1-\mathcal{P}_m qf]V(j,k \mid m+1)]u \\ &+(1-u)V(j+1,k+1 \mid 0)\}, \\ k=0,\ldots,(M-1), &j=1,\ldots,(N-1). \end{aligned} \tag{38}$$

If the DM is alive and at the last target (which may still be alive) and has $k>0$ weapons at hand, it is clearly optimal to deploy a weapon since there is no payoff in retaining weapons. Indeed, for any $k,\, m \geq 0$:

$$\begin{aligned} V(N,k+1 \,|\, m) = &\, p\mathcal{P}_m r_N + \mathcal{P}_m qfsV(N,k \,|\, 0) \\ &+(1-\mathcal{P}_m qf)V(N,k \,|\, m+1), \end{aligned} \tag{39}$$

with the boundary condition $V(N,\, 0 \,|\, m) = 0$ for any $m \geq 0$.

As before, we are interested in determining if the optimal policy is threshold based and if it has desirable monotonicity properties. Motivated by the perfect information result, we define the corresponding (equivalent) partial information marginal future reward:

$$\Delta_j(k) \coloneqq V(j,k+1|0) - (p+qa)V(j,k|0), \quad \forall j, \tag{40}$$

where $a = 1-f(1-s)$. For the case of $f=1$ (perfect information), we have shown that the marginal future reward is the threshold at which the bomber's control switches and furthermore, the threshold exhibits desirable monotonicity properties in both k and j. Unfortunately, for the partial information setup, a similar analysis does not work. In other words, we show that the marginal future reward (40) is *not* the threshold at which the bomber's control switches. We do so via a counterexample for the special case of an invincible bomber, that is, $s=1$.

3.2 | Invincible bomber counterexample

For the special case $s=1$, the partial information Bellman recursion (37) is given by:

$$\begin{aligned} V(j,k+1|m) = \max_{u=0,1}\{&[p\mathcal{P}_m r_j + \mathcal{P}_m qfV(j,k|0) \\ &+(1-\mathcal{P}_m qf)V(j,k|m+1)]u \\ &+(1-u)V(j+1,k+1|0)\}. \end{aligned} \tag{41}$$

Repeating the steps (6)–(9) in Proposition 1, one can show that the corresponding optimal policy satisfies:

$$\mu(j,k+1|m) = 1, \quad p\mathcal{P}_m r_j \geq \Delta_{j+1}(k). \tag{42}$$

Note that for $s = 1$, $\Delta_{j+1}(k) = V(j+1, k+1|0) - V(j+1, k|0)$ and is therefore identical to the perfect information marginal reward defined earlier (5). So, it is not unreasonable to expect the marginal future reward to be the threshold at which the bomber's control switches. Indeed, we have the partial result in (42) that gives us a sufficient condition for bombing the current target. Unfortunately, it is no longer the case that $\mu(j, k+1|m) = 0$, if $p\mathcal{P}_m r_j < \Delta_{j+1}(k)$. Suppose we have two targets, that is, $N = 2$ and the DM is at the first target with $m = k = 1$. We will show that there exists r_1 for which it is optimal to bomb the current target even though the immediate expected payoff is less than the marginal future reward (40). In other words, $\mu(1, 2|1) = 0$ even though $p\mathcal{P}_1 r_1 < \Delta_2(1)$.

Lemma 1

$$p\mathcal{P}_1 r_1 < \Delta_2(1) \;\; \textit{but} \;\; \mu(1,2|1) = 0, \;\; \textit{if} \;\; r_1 \in (L_1 r_2, L_2 r_2),$$
$$\textit{where:} \quad L_1 = \frac{1 - qf^2}{1 - pf - qf^2} \quad \textit{and} \quad L_2 = \frac{1 - qf}{1 - f}. \tag{43}$$

Proof First we note that it can easily be shown that $1 < L_1 < L_2$ so we can always pick r_1 from the nonempty interval $(L_1 r_2, L_2 r_2)$. From the a posteriori probability update (34), we have:

$$\mathcal{P}_1 = \frac{q(1-f)}{1 - qf} = \frac{q}{L_2}. \tag{44}$$
$$\Rightarrow \frac{q(1 + \mathcal{P}_1 f)}{\mathcal{P}_1(1 + qf)} = \frac{1 - qf^2}{1 - pf - qf^2} = L_1. \tag{45}$$

From the Bellman recursion (41), we can write:

$$V(2, k|m) = p\mathcal{P}_m r_2 + \mathcal{P}_m qf V(2, k-1|0) + (1 - \mathcal{P}_m qf) V(2, k-1|m+1) \tag{46}$$
$$\Rightarrow V(2, 1|0) = pr_2. \tag{47}$$

By repeated application of (46), we have:

$$\begin{aligned} V(2,2|0) &= pr_2 + qfV(2,1|0) + (1 - qf)V(2,1|1) \\ &= pr_2 + qfpr_2 + (1 - qf)p\mathcal{P}_1 r_2 \\ &= pr_2 + qfpr_2 + pq(1-f)r_2 = pr_2(1+q). \end{aligned} \tag{48}$$

So, we can write:

$$\Delta_2(1) = V(2,2|0) - V(2,1|0) = pqr_2. \tag{49}$$

Since $r_1 < L_2 r_2$, we can write:

$$p\mathcal{P}_1 r_1 < p\mathcal{P}_1 L_2 r_2 = pqr_2 = \Delta_2(1). \tag{50}$$

From the Bellman recursion (41), we have:

$$V(1,1|0) = \max_{u=0,1}\{pr_1 u + (1-u)V(2,1|0)\} = pr_1. \tag{51}$$
$$V(1,1|2) = \max_{u=0,1}\{p\mathcal{P}_2 r_1 u + (1-u)V(2,1|0)\} = pr_2. \tag{52}$$

$$\begin{aligned} V(1,2|1) &= \max_{u=0,1}\{[p\mathcal{P}_1 r_1 + q\mathcal{P}_1 f V(1,1|0) \\ &\quad + (1 - q\mathcal{P}_1 f)V(1,12)]u \\ &\quad + (1-u)V(2,2|0)\} \\ \Rightarrow \mu(1,2|1) &= \arg\max_{u=0,1}\{p\mathcal{P}_1 r_1 u \\ &\quad + q\mathcal{P}_1 f[V(1,1|0) - V(2,1|0)]u + (1-u)\Delta_2(1)\} \\ &= \arg\max_{u=0,1}\{p\mathcal{P}_1 r_1 u + pq\mathcal{P}_1 f[r_1 - r_2]u \\ &\quad + (1-u)pqr_2\} \\ &= \arg\max_{u=0,1}\left\{ r_1 u + \frac{q(1+\mathcal{P}_1 f)}{\mathcal{P}_1(1+qf)} r_2(1-u) \right\} \\ &= \arg\max_{u=0,1}\{r_1 u + (1-u)L_1 r_2\} = 1. \end{aligned} \tag{53}$$

Note that (52) follows from (50) since $\mathcal{P}_2 < \mathcal{P}_1$ and so, $p\mathcal{P}_2 r_1 < p\mathcal{P}_1 r_1 < pqr_2 < pr_2$. ∎

So, we conclude that for the partial information scenario, a thresholding policy *similar to the perfect information case* cannot be established. We emphasize that we have only shown that the marginal future reward (40) is *not* the threshold at which the optimal policy switches. It is quite likely that the optimal bombing policy is a thresholding policy with desirable monotonicity properties but we do not have a definitive proof either way at this time. However, in light of the negative result established herein, we propose a threshold based heuristic for the partial information case.

3.3 | Threshold based heuristic

Motivated by the perfect information threshold policy, we propose the following threshold based heuristic policy for the partial information scenario. For any $j < N$, let $\widetilde{\kappa}(j, m)$ be the smallest $k = 0, 1, \ldots$ such that $p\mathcal{P}_m r_j \geq \widetilde{\Delta}_{j+1}(k)$ for all $j = 1, \ldots, (N-1)$, where the approximate marginal reward $\widetilde{\Delta}_j(k)$ is computed according to:

$$\widetilde{\Delta}_j(k) = \begin{cases} \widetilde{\Delta}_{j+1}(k), & k < \widetilde{\kappa}(j,0), \\ pr_j, & k = \widetilde{\kappa}(j,0), \\ p\sum_{i=0}^{\ell-1}(qa)^i \widetilde{\Delta}_{j+1}(k-i-1) & \\ \quad + pr_j\left[(qa)^\ell + q(1-a)\sum_{i=0}^{\ell-1}(qa)^i\right], & \\ \ell = k - \widetilde{\kappa}(j,0) > 0. & \end{cases} \tag{54}$$

The boundary condition is given by:

$$\widetilde{\Delta}_N(k) = \frac{pr_N}{1 - qa}[q(1-a) + p(qa)^k]. \tag{55}$$

The threshold based heuristic policy for all $j = 1, \ldots, (N-1)$ is prescribed by:

$$\widetilde{\mu}(j, k+1|m) = \begin{cases} 0, & k < \widetilde{\kappa}(j,m), \\ 1, & \text{otherwise.} \end{cases} \tag{56}$$

At the last (possibly alive) target, it is always optimal to bomb: $\widetilde{\mu}(N, k+1|m) = 1, k, m \geq 0$. Note that Equations (54)–(56) closely mirror the perfect information Equations (30)–(32) and they coincide when $f = 1$. Indeed, for $f = 1 \Rightarrow a = s$ and the heuristic policy prescribed above is optimal. Furthermore,

the approximate marginal reward function defined herein is not arbitrary in that it equals the exact marginal reward at least for the last target. In other words,

$$\widetilde{\Delta}_N(k) = \Delta_N(k) = V(N, k+1|0) - (p+qa)V(N, k|0). \quad (57)$$

Lemma 2

$$V(N, k|0) = pr_N \sum_{\ell=0}^{k-1} (qa)^\ell. \quad (58)$$

Proof We shall prove the result by induction on k. From (39), we immediately have: $V(N, 1|0) = pr_N$. Assume that for any t; $1 \le t < k$,

$$V(N, t \mid 0) = pr_N \sum_{\ell=0}^{t-1} (qa)^\ell. \quad (59)$$

We proceed to show that:

$$V(N, k|0) = pr_N \sum_{\ell=0}^{k-1} (qa)^\ell. \quad (60)$$

Indeed, we have from the Bellman recursion (39) applied to the last target:

$$\begin{aligned} V(N,k|0) &= pr_N + qfsV(N,k-1|0) + (1-qf)V(N,k-1|1) \\ &= pr_N + qfs\left[pr_N \sum_{\ell=0}^{k-2}(qa)^\ell\right] + (1-qf) \\ &\quad \times [p\mathcal{P}_1 r_N + \mathcal{P}_1 qfsV(N,k-2|0) \\ &\quad + (1-\mathcal{P}_1 qf)V(N,k-2|2)] \quad (61) \\ &= pr_N\left[1 + qfs\sum_{\ell=0}^{k-2}(qa)^\ell\right] \\ &\quad + q(1-f)[pr_N + qfsV(N,k-2|0)] \\ &\quad + (1-qf)(1-\mathcal{P}_1 qf)V(N,k-2|2) \quad (62) \\ &= pr_N\left[1 + qfs\sum_{\ell=0}^{k-2}(qa)^\ell\right] \\ &\quad + q(1-f)pr_N\left[1 + qfs\sum_{\ell=0}^{k-3}(qa)^\ell\right] \\ &\quad + (1-qf)(1-\mathcal{P}_1 qf)V(N,k-2|2) \quad (63) \\ &= pr_N\left[1 + qfs\sum_{\ell=0}^{k-2}(qa)^\ell\right] \\ &\quad + q(1-f)pr_N\left[1 + qfs\sum_{\ell=0}^{k-3}(qa)^\ell\right] \\ &\quad + q^2(1-f)^2 pr_N\left[1 + qfs\sum_{\ell=0}^{k-4}(qa)^\ell\right] \\ &\quad + \cdots + q^{k-1}(1-f)^{k-1}pr_N \quad (64) \\ &= pr_N\sum_{m=2}^{k} q^{k-m}(1-f)^{k-m}\left[1 + qfs\sum_{\ell=0}^{m-2}(qa)^\ell\right] \\ &\quad + pr_N q^{k-1}(1-f)^{k-1}. \quad (65) \end{aligned}$$

In the above derivation, we have repeatedly used the induction assumption (59) for $t \le k-1$. We have also used the probability update Equation (34):

$$\begin{aligned} &\mathcal{P}_m = \frac{\mathcal{P}_{m-1}q(1-f)}{1-\mathcal{P}_{m-1}qf}, \quad m > 0 \quad \text{and} \quad \mathcal{P}_0 = 1. \\ &\Rightarrow \mathcal{P}_m \prod_{i=0}^{m-1}[1-\mathcal{P}_i qf] = q^m(1-f)^m, \quad m > 0. \quad (66) \end{aligned}$$

Finally, we have used the boundary condition: $V(N, 1|k-1) = p\mathcal{P}_{k-1}r_N$ to arrive at (65). Note however that, $1-f+fs = 1-f(1-s) = a$ and so, we can write:

$$\begin{aligned} \sum_{\ell=0}^{k-1}(qa)^\ell &= 1 + \sum_{\ell=1}^{k-1}(qa)^\ell = 1 + q\sum_{\ell=1}^{k-1}(qa)^{\ell-1}(1-f+fs) \\ &= 1 + q\left[(1-f)\sum_{\ell=0}^{k-2}(qa)^\ell + fs\sum_{\ell=0}^{k-2}(qa)^\ell\right] \\ &= \left[1 + qfs\sum_{\ell=0}^{k-2}(qa)^\ell\right] + q(1-f)\underbrace{\sum_{\ell=0}^{k-2}(qa)^\ell}. \quad (67) \end{aligned}$$

By repeated application of (67) to the underbraced expression, we get:

$$\begin{aligned} \sum_{\ell=0}^{k-1}(qa)^\ell &= \sum_{m=2}^{k} q^{k-m}(1-f)^{k-m}\left[1 + qfs\sum_{\ell=0}^{m-2}(qa)^\ell\right] \\ &\quad + q^{k-1}(1-f)^{k-1} \quad (68) \end{aligned}$$

Upon comparing (65) and (68), we conclude that:

$$V(N, k|0) = pr_N \sum_{\ell=0}^{k-1}(qa)^\ell. \quad (69)$$

■

It immediately follows that the marginal reward for the last target:

$$\begin{aligned} \Delta_N(k) &= V(N,k+1|0) - (p+qa)V(N,k|0) \\ &= \frac{pr_N}{1-qa}\{1 - (qa)^{k+1} - [p+qa][1-(qa)^k]\} \\ &= \frac{pr_N}{1-qa}\{1 - p[1-(qa)^k] - qa\} \\ &= \frac{pr_N}{1-qa}[q(1-a) + p(qa)^k], \quad (70) \end{aligned}$$

which by definition equals the approximate marginal future reward, $\widetilde{\Delta}_N(k)$. To summarize, we have shown via a counterexample that the threshold based heuristic policy is sub-optimal for $f < 1$. We have also shown that it is optimal when $f = 1$. In the absence of theoretical bounds on the quality of the heuristic policy, we perform a numerical study instead. Towards this end, we first compute the value yielded by implementing the heuristic policy. Let $\widetilde{V}(j, k+1|m)$ denote the value yielded to the DM when it implements the threshold

heuristic. Accordingly, we have:

$$\begin{aligned}\widetilde{V}(j,k+1|m) = [&p\mathcal{P}_m r_j + \mathcal{P}_m q f s \widetilde{V}(j,k \mid 0) \\ &+ (1-\mathcal{P}_m q f)\widetilde{V}(j,k|m+1)]\widetilde{\mu}(j,k+1|m) \\ &+ [1-\widetilde{\mu}(j,k+1 \mid m)]\widetilde{V}(j+1,k+1 \mid 0), \\ k = 0,\ldots,(M-1), \quad &j = 1,\ldots,(N-1). \end{aligned} \tag{71}$$

Since the heuristic policy for the last (Nth) target coincides with the optimal policy, we also have:

$$\widetilde{V}(N,k+1|m) = V(N,k+1|m), \quad \forall k,m \geq 0.$$

For different M and N values, we perform Monte-Carlo simulations to provide empirical bounds on the loss in total expected reward at the beginning of the bombing mission, that is, $V(1,M|0) - \widetilde{V}(1,M|0)$. In addition, we also compute the percent difference in the optimal and heuristic policies over all reachable states starting from the initial state: $j = 1, k = M$, $m = 0$. Figure 2 shows box plot analyses of the empirical loss in value and difference from optimal policy for four different sets of M and N values. For each set, we compute statistics for 10 different values of $f \in [0, 1]$. All other problem parameters are randomly sampled from uniform distributions: $p \in [0, 1]$, $s \in [0, 1]$ and $r_j \in [1, \ldots, 50]$ for all $j = 1, \ldots, N$. For each (M, N) value, we perform 1,000 Monte-Carlo runs. From Figure 2, it appears that the loss in value increases as the number of targets N increases. A similar trend holds for the percent difference between the optimal and heuristic policies. We stop at $N = 8$, $M = 12$ since the computation time for solving the exact dynamic program grows exponentially. As expected, there is no loss in optimality for all data sets with $f = 1$. Interestingly, we also see that the performance degradation starts low for small f, grows and then decreases to zero at $f = 1$. Needless to say, these observations are based on simulation results and we have no conclusive proof of this property at the time of writing this article. However, an encouraging sign is the relatively small loss in performance incurred in using the heuristic compared to the potential savings in computation time achieved in computing the heuristic policy (via a direct linear recursion).

4 | PARTIAL INFORMATION GAME

In this section, we consider the following sequence of events. Suppose the bomber fires at a live target. If the target survives the attack, it can either return fire or play dead. The bomber waits (looks) to see if there is return fire before proceeding with the next action. Given that a round of return fire has a positive probability of destroying the bomber, it would seem that the target should always return fire, when presented with an opportunity. Conversely, when the targets are cooperating, it may be advantageous for a high value target to sometimes play dead so that the bomber moves on to a low value target downstream.

As in the counter example earlier, we consider a representational two target scenario that captures the essence of the game. Indeed, let the bomber be at the first target T_1 with two weapons at hand. Further suppose that $r_1 > r_2$ (else the bomber may simply move on to T_2) and the bomber fires the first shot at T_1 and is waiting to see if there is any return fire. We wish to answer the following key questions. Should T_1, if it survives the first round of direct fire, return fire or play dead? Should the bomber, if still alive, expend the last weapon at the first or the second target? We investigate this setup and compute the Nash equilibrium strategies for both players.

Since we are dealing with a partial information game, the decisions made by the players are a function of their respective information states. A player's strategy is a mapping from his information state to available actions. We first consider the strategy of the bomber, who strives to maximize the reward yielded by inflicting damage on T_1 and/or T_2. If it sees return fire and survives it, the bomber knows that T_1 is alive and expends the last weapon on it since $r_1 > r_2$. This part is straightforward. However, if there is no return fire, it is not clear what the Bomber's strategy should be. For this instance, suppose that the bomber's strategy is to fire the second shot at T_1 with probability b. Furthermore, if T_1 survives the first round of fire, let the probability with which it returns fire be given by f. We will compute the optimal values for b and f that is, the Nash equilibrium strategies. We will also illustrate conditions under which pure strategies are optimal.

Let (b, f) indicate the tuple of mixed strategies for the bomber and T_1. Suppose T_1 survived the first round of fire. The expected damage as *seen* by T_1 for the joint strategies (b, f) is given by:

$$V_T(b,f) = fspr_1 + (1-f)[bpr_1 + (1-b)pr_2]. \tag{72}$$

In other words, if T_1 fires back, it knows that the bomber will survive the return fire with probability s and fire at it again yielding an expected damage of pr_1. On the other hand, if it holds fire (plays dead), the bomber's strategy will yield an expected damage of $bpr_1 + (1-b)pr_2$. For the same pair of strategies, the expected reward yielded by the second round of fire as *seen* by the bomber is given by:

$$V_B(b,f) = p[b \times 0 + (1-b)pr_2] + qV_T(b,f). \tag{73}$$

Indeed, if the first round was successful, the bomber gets the reward $(1-b)pr_2$ from the second round. On the other hand, if the first round was unsuccessful, it gets the reward $V_T(b,f)$ from the second round of fire. The target T_1 strives to minimize the damage $V_T(b,f)$ and the bomber strives to maximize the reward $V_B(b,f)$. We shall first compute each player's best response to a given opponent strategy. For a given bomber strategy, b, T_1's best response, $f^*(b)$ that minimizes $V_T(b,f)$ is computed as follows. Recall that:

$$\begin{aligned} V_T(b,f) = p[&(s-b)r_1 - (1-b)r_2]f \\ &+ bpr_1 + (1-b)pr_2. \end{aligned} \tag{74}$$

By inspection, it follows that:

$$f^*(b) = \begin{cases} 0, & \text{if } (s-b)r_1 > (1-b)r_2, \\ 1, & \text{if } (s-b)r_1 < (1-b)r_2. \end{cases} \tag{75}$$

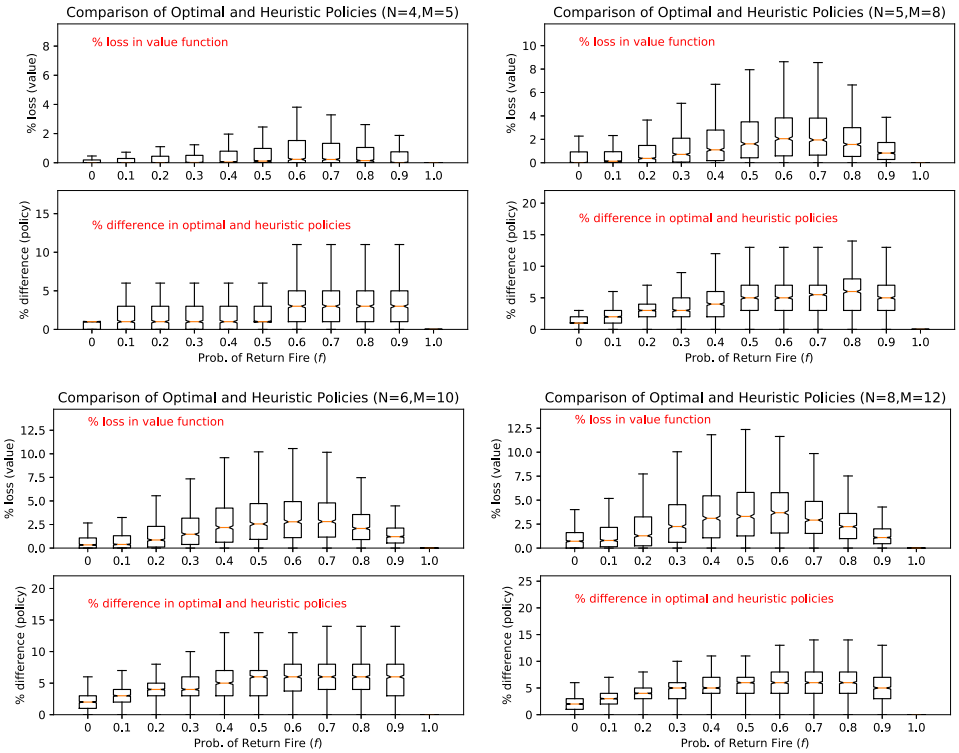

FIGURE 2 Percentage loss in optimal value and difference from optimal policy for different problem parameters [Colour figure can be viewed at wileyonlinelibrary.com]

In particular, if $b = \frac{sr_1 - r_2}{r_1 - r_2}$, $V_T(b,f)$ is independent of f and hence, any value $f \in [0, 1]$ is optimal. In a similar fashion, for a given T_1 strategy f, the bomber's best response, $b^*(f)$ that maximizes $V_B(b,f)$ is computed as follows. Recall that:

$$\begin{aligned} V_B(b,f) &= p(1-b)pr_2 + q\{fspr_1 + (1-f)[bpr_1 + (1-b)pr_2]\}. \\ &= p[q(1-f)r_1 - (1-fq)r_2]b \\ &\quad + qfspr_1 + (1-fq)pr_2 \end{aligned} \tag{76}$$

By inspection, it follows that:

$$b^*(f) = \begin{cases} 1, & \text{if } q(1-f)r_1 > (1-fq)r_2, \\ 0, & \text{if } q(1-f)r_1 < (1-fq)r_2, \end{cases} \tag{77}$$

In particular, if $f = \frac{qr_1 - r_2}{q(r_1 - r_2)}$, $V_B(b,f)$ is independent of b and hence, any value $b \in [0, 1]$ is optimal. By definition, the Nash equilibrium strategies, $(\widetilde{b},\widetilde{f})$ satisfy $\widetilde{b} = b^*(\widetilde{f})$ and $\widetilde{f} = f^*(\widetilde{b})$. We have the following result establishing the Nash equilibrium.

Theorem 3

$$(\widetilde{b},\widetilde{f}) = \begin{cases} (0,1), & \textit{if} \quad sr_1 < r_2, \\ (0,0), & \textit{if} \quad qr_1 < r_2 < sr_1, \\ \left(\frac{sr_1 - r_2}{r_1 - r_2}, \frac{qr_1 - r_2}{q(r_1 - r_2)}\right), & \textit{otherwise}. \end{cases} \tag{78}$$

Proof We shall prove the result by considering the different inequalities:

Case 1. $sr_1 < r_2$: It follows from (74) that $V_T(0, f) = p(sr_1 - r_2)f + pr_2$ and therefore, the minimizing best target response, $f^*(0) = 1$. Also, from (76), we have $V_B(b, 1) = -p^2r_2b + qspr_1 + p^2r_2$ and so, the maximizing best bomber response, $b^*(1) = 0$. Hence, $(\widetilde{b},\widetilde{f}) = (0, 1)$ are the Nash equilibrium pure strategies that is, T_1 always returns fire and the bomber always fires at T_2 in the event of no return fire. The corresponding expected damage seen by T_1 and reward seen by the bomber are given by $V_T(0, 1) = psr_1$ and $V_B(0, 1) = p^2r_2 + qpsr_1$ respectively.

Case 2. $qr_1 < r_2 < sr_1$: It follows from (74) that $V_T(0, f) = p(sr_1 - r_2)f + pr_2$ and therefore, the minimizing best target response, $f^*(0) = 0$. Also, from (76), we have $V_B(b, 0) = p(qr_1 - r_2)b + pr_2$ and so, the maximizing best bomber response, $b^*(0) = 0$. Hence, $(\widetilde{b},\widetilde{f}) = (0, 0)$ are the Nash equilibrium pure strategies that is, T_1 always plays dead and the bomber always fires at T_2 in the event of no return fire. The corresponding expected damage seen by T_1 and reward seen by the bomber are given by $V_T(0, 0) = pr_2$ and $V_B(0, 0) = pr_2$ respectively.

Case 3. $r_2 \le \min(qr_1, sr_1)$: For this case, no pure strategies exist. Indeed, the game does not admit the other possible solutions, (1, 1) and (1, 0). From (74), we have $V_T(1, f) = p(s-1)r_1f + pr_1$ and therefore, the minimizing best target response, $f^*(1) = 1$ since $s < 1$. However, from (76), we have $V_B(b,$

1) $= -p^2r_2b + qspr_1 + p^2r_2$ and so, the maximizing best bomber response, $b^*(1) = 0$. So, the bomber switches strategy. Continuing in this fashion, we note that $V_T(0, f) = p(sr_1 - r_2)f + pr_2$. Since, $r_2 < sr_1$, the target also switches strategy to $f^*(0) = 0$. Finally, we note that $V_B(b, 0) = p(qr_1 - r_2)b + pr_2$ and so, the maximizing best bomber response, $b^*(0) = 1$ since $qr_1 > r_2$. So, the bomber switches strategy again implying that there are no Nash equilibrium pure strategies for this case. So, the only admissible solution is the tuple of mixed strategies, $(\widetilde{b},\widetilde{f}) = \left(\frac{sr_1 - r_2}{r_1 - r_2}, \frac{qr_1 - r_2}{q(r_1 - r_2)}\right)$ that simultaneously makes $V_B(b,\widetilde{f})$ to be independent of b and $V_T(\widetilde{b},f)$ to be independent of f. Note that the mixed strategy probabilities are well defined that is, $\widetilde{b},\widetilde{f} \in [0, 1)$ since $r_2 \leq sr_1 < r_1$ and $qr_2 < r_2 \leq qr_1$. The corresponding expected damage seen by T_1 and reward seen by the bomber are given by:

$$V_T(\widetilde{b},\widetilde{f}) = p(r_1 - r_2)\widetilde{b} + pr_2 = psr_1. \tag{79}$$

$$V_B(\widetilde{b},\widetilde{f}) = qp(sr_1 - r_2)\widetilde{f} + pr_2$$
$$= p(qr_1 - r_2)\widetilde{b} + pr_2 = \frac{psr_1(qr_1 - r_2) + p^2r_1r_2}{r_1 - r_2}. \tag{80}$$ ■

In this article, we do not consider the generalization of the above result to the case of N targets and M weapons. Suffice to say that this will be a challenging enterprise given the interplay between the information states for the two players that is, the bomber and the target team. It is reasonable to expect that the optimal play for the jth target will rely on a tradeoff between the expected personal damage, if it returns fire, and expected marginal damage to downstream targets, if it chooses to play dead. The optimal play also relies on what the bomber's expectation of its rewards are. Note that since we are dealing with a partial information scenario, the expected rewards calculated by the two players are different that is, this is not a zero sum game. This makes it a formidable challenge to solve the general case. A possible line of attack would be to employ the dynamic programming approach in Sakaguchi (1977), which deals with a sequential engagement with perfect information, where the attacker/defender choose to attack/defend each target by firing a single shot.

5 | CONCLUSIONS AND FUTURE WORK

We consider a dynamic variant of the weapon-target assignment problem, wherein ground targets are sequentially visited by a bomber equipped with homogenous weapons. We investigate the scenario where the targets are capable of retaliation and solve for the optimal play therein. In particular, we consider an interesting informational aspect of the game, where the act of firing back at the bomber reveals the status of the target. To complete the analyses, we also consider the cases where the target action is deterministic and random. Insightful monotonicity properties of the threshold function at which the bomber's control switches are preserved under perfect information. Future work will focus on either establishing a monotonic threshold policy or providing counterexamples thereof for the partial information setup. Finally, for a two target game scenario, we develop Nash equilibrium mixed strategies and establish conditions under which pure strategies are optimal. In the future, we plan to extend this result to the general case of multiple targets and multiple homogenous weapons with the bomber.

ORCID

Krishna Kalyanam https://orcid.org/0000-0003-4670-7012

REFERENCES

Aviv, Y., & Kress, M. (1997). Evaluating the effectiveness of shoot-look-shoot tactics in the presence of incomplete damage information. *Military Operations Research*, *3*(1), 79–89.

Aydin, S., Akçay, Y., & Karaesmen, F. (2009). On the structural properties of a discrete-time single product revenue management problem. *Operations Research Letters*, *37*(4), 273–279.

Gao, X., Zhong, W., & Mei, S. (2013). A game-theory approach to configuration of detection software with decision errors. *Reliability Engineering and System Safety*, *119*, 35–43.

Glazebrook, K. D., Mitchell, H. M., Gaver, D. P., & Jacobs, P. A. (2004). The analysis of shooting problems via generalized bandits. Tech. Rep. NPS-OR-04-005, Naval Postgraduate School, Monterey, CA.

Glazebrook, K. D., & Washburn, A. (2004). Shoot-look-shoot: A review and extension. *Operations Research*, *52*(3), 454–463.

Hu, X., Xu, M., Xu, S., & Zhao, P. (2017). Multiple cyber attacks against a target with observation errors and dependent outcomes: Characterization and optimization. *Reliability Engineering and System Safety*, *159*, 119–133.

Kalyanam, K., Casbeer, D., & Pachter, M. (2017). Monotone optimal threshold feedback policy for sequential weapon target assignment. *AIAA Journal of Aerospace Information Systems*, *14*(1), 68–72.

Kalyanam, K., Casbeer, D., & Pachter, M. (2018). Optimal sequential resource allocation under error-prone success assessment. In *10th IMA International Conference on Modelling in Maintenance and Reliability*, Manchester, U.K. (pp. 66–72).

Kalyanam, K., Rathinam, S., Casbeer, D., & Pachter, M. (2016). Optimal threshold policy for sequential weapon target assignment. In *IFAC Symposium on Automatic Control in Aerospace*, Vol. 49 of *IFAC-PapersOnLine*, Sherbrooke, Quebec, Canada (pp. 7–10).

Kamihigashi, T. (2017). Counterexamples to property B of the discrete time bomber problem. *Annals of Operations Research*, *248*(1–2), 579–588.

Kisi, T. (1976). Suboptimal decision rule for attacking targets of opportunity. *Naval Research Logistics*, *23*(3), 525–533.

Levitin, G., & Hausken, K. (2012). Resource distribution in multiple attacks with imperfect detection of the attack outcome. *Risk Analysis*, *32*(2), 304–318.

Mastran, D. V., & Thomas, C. J. (1973). Decision rules for attacking targets of opportunity. *Naval Research Logistics*, *20*(4), 661–672.

Sakaguchi, M. (1977). A sequential allocation game for targets with varying values. *Journal of the Operations Research Society of Japan*, *20*(3), 182–193.

Sato, M. (1996). A sequential allocation problem with search cost where the shoot-look-shoot policy is employed. *Journal of the Operations Research Society of Japan*, *39*(3), 435–454.

Sato, M. (1997a). On optimal ammunition usage when hunting fleeing targets. *Probability in the Engineering and Informational Sciences*, *11*, 49–64.

Sato, M. (1997b). A stochastic sequential allocation problem where the resources can be replenished. *Journal of the Operations Research Society of Japan*, *40*(2), 206–219.

van Ryzin, G. J., & Talluri, K. T. (2005). *An introduction to revenue management*. In INFORMS TutORials in Operations Research (pp. 142–194). Berlin: Springer.

Weber, R. R. (2013). ABCs of the bomber problem and its relatives. *Annals of Operations Research*, *208*(1), 187–208.

How to cite this article: Kalyanam K, Casbeer D, Pachter M. A sequential partial information bomber-defender shooting problem. *Naval Research Logistics* 2020;67:223–235. https://doi.org/10.1002/nav.21892