Proceedings of the 9th World Congress on Electrical Engineering and Computer Systems and Sciences (EECSS'23)
Brunel University, London, United Kingdom - August 03-05, 2023
Paper No. MVML 103
DOI: 10.11159/mvml23.103

Dempster-Shafer Evidential Theory Belief Amalgamation and Dynamic Programming Supporting Soldier Squadron Adversarial Engagement: Simulation-Based Decision-Making

Nicholas V. Scott[1], Bradon Thymes[2], and Joseph P. Salisbury[3]
Riverside Research Institute[1], Open Innovation Center, Dayton Research Center,
2640 Hibiscus Way, Beavercreek, OH, 45431, USA
nscott@riversideresearch.org
Cornell University[2], Department of Computer Science,
402 Gates Hall, Ithaca, NY 14853 Dayton Research Center
bthymes@cornell.edu
Riverside Research Institute[3], Open Innovation Center, Boston Research Office
70 Westview Street, Lexington, MA, 02421, USA
jsalisbury@riversideresearch.org

Abstract

Increasing complexity of adversarial engagement in urban warfare scenarios suggests that there is a serious demand for technologies that can assist the warfighter's decision-making capabilities. This is especially true for stressful conditions experienced by soldiers where optimal and expedient execution of missions and directives is needed under the yoke of resource constraint. In such situations, the age-old military and historical hallmarks of squadron group belief consensus supporting mission accomplishment and optimal movement of a unit based on probabilistic assessment of value come to the fore and are vital elements in supporting and preserving mission integrity. Both fundamental drivers for squadron success can be supported and demonstrated using a two-element statistical machine learning formulism. The first part utilizes Dempster-Shafer evidential theory for belief amalgamation across a squadron and the second part dynamic programming for estimation of optimal path movement through a hostile domain. Simulations using these two important geo-intelligence processors are performed to demonstrate machine learning-based processing which can ultimately be utilized in intelligent squad weapon systems aiding in decision-making.

The first formulism is developed around the geo-intelligence scenario of a 4-soldier squadron trying to decide which direction to proceed based on the amalgamation of individual beliefs. Dempster-Shafer evidential theory provides a way to amalgamate the beliefs of each soldier via a method that uses orthogonal summation of probability mass associated with different propositions. Eleven propositions exist for the 4-directional problem encompassing not only the initial 4 directions, but propositions based on logical 'or' as well as the state of ignorance parameterizing uncertainty. An orthogonal amalgamation template allows for amalgamation of belief information for soldiers 1 and 2 and then soldiers 3 and 4. These soldier beliefs are then amalgamated into a virtual soldier for all soldiers representing the squadron group mind. Simulation results demonstrate that given the initial probability mass profiles for each soldier, direction 2 is the optimal direction to proceed in providing a logical guide for squadron movement.

The increasing complexity of warfighter scenarios with respect to adversarial engagement and non-traditional environments of contention also suggests a serious need to leverage technology to go beyond mere state estimation towards algorithms prescribing actions. Dynamic programming is well suited for policy-based decision-making where there is a need to assist humans in making decisions where the best choices or actions are not clear and depend on *values* placed on specific situations or states. To demonstrate the applicability and power of dynamic programming to path-based decision-making, a fictious model problem of finding the optimal policy for moving a soldier squadron through a multiple room building is addressed. The objective is to illustrate how prior geo-intelligence information in the form of transition probabilities and rewards can be used to facilitate decision-making in terms of what *should* optimally be done rather than what must be done. The objective is also focused on understanding the effect of noise infiltration on optimal policy estimation.

A soldier squadron is tasked with moving through a 4-floor building and 'clearing' it. A 16 state-2 action dynamic programming model for building 'clearing' is created where the aim is to understand on *average* what rooms should be places of hostile engagement (action 1) and what rooms should not (action 2). Prior ground clandestine geo-intelligence provides the room-to-room transition probability and reward fields. The geo-intelligence dynamic programming issue of interest is *how* the optimal policy (function mapping of state to action) changes as noise infiltrates the geo-intelligence system, changing the transition probability and reward fields.

Baseline transition probability and reward fields for the 16 state-2 action domain are set up where particular attention is paid to their prescription based on underlying fire fight capability. The optimal policy resulting from the baseline transition probability and reward fields demonstrates that hostile engagement is optimal *along* a floor while passivity is optimal at the floor transition points. This is consistent with floor transitions being choke points where quick movement rather than fighting is more highly rewarded. When the reward field for hostile engagement is lowered near the end room floor transition points, continual hostile engagement completely along a floor is no longer optimal. Hostile engagements should be performed near the beginning and middle of the floor.

Noise infiltrating prior clandestine geo-intelligence causes modulation of the state transition probability field which cause changes in the baseline optimal policy. Results suggests that uncertainty in the transition probability field associated with one type of action precipitates increases in the alternate action. This suggests that increases in hostile engagement actions are associated with increased uncertainty in transition probabilistic information for the non-hostile engagement action. Increases in non-hostile engagement action are associated with increased uncertainty in transition probabilistic information for the hostile-engagement action.

Keywords: action, belief, building, Dempster-Shafer evidential theory, dynamic programming, floors, hostile engagement, non-hostile engagement, optimal policy, propositions, reward field, rooms, state, transition probability field

1. Introduction

With increasing complexity of warfighter scenarios with respect to adversarial engagement and non-traditional environments of contention, there is an increasing need to leverage technology to process geo-intelligence information to develop not only optimal state insight but also optimal decision-making. Dempster-Shafer (D-S) evidential theory-based belief amalgamation [1] is well suited to information fusion needed in different geo-intelligence scenarios including group squadron optimal direction estimation for adversarial engagements. The fictious model problem of deciding which direction a squadron should proceed as a unit after a fire fight engagement is addressed to demonstrate the power of D-S evidential theory for group belief amalgamation.

Similar warfighter scenarios suggest a serious need to leverage technology to go beyond mere state estimation towards algorithms that prescribe actions. Dynamic programming [2] is well suited for policy-based decision-making where there is a need to assist humans in making decisions. In these situations, the best choices or actions are not absolutely and deterministically clear and depend on values placed on specific situations or states. The applicability of dynamic programming to path-based decision-making is demonstrated via the formulation and solution of the fictious model problem of finding the optimal policy for moving a squadron through a multiple room building. The objective is to illustrate how previous geo-intelligence information in the form of transition probabilities and rewards (costs) can be used to facilitate decision making in terms of what *should* optimally be done rather than what *must* be done. Such evaluation of state and action is useful in facilitating what optimal pathways should be adhered to, allowing for the mitigation of mistakes that can cost human life.

2. Problem Scenario and Methodology for Soldier Squadron Belief Amalgamation

The problem scenario addressed using D-S evidential theory is as follows. A 4-soldier squadron has just finished a fire-fight engagement where artillery fire from four different directions have been received. Machine learning (ML) aided assault rifle technology has provided the state estimate that the direction of highest muzzle flash variance and therefore the most appropriate direction to proceed in is direction 2 from a range of four different direction choices. (The four directions span four equisector divisions of a 180-degree angular domain situated in front of the squadron). This machine learning estimate is communicated to each individual soldier and facilitates cognitive construction of individual beliefs as to which direction is optimal. There is a desire to amalgamate the beliefs of all soldiers to provide a fuller understanding of which direction is optimal. D-S evidential theory provides a way to amalgamate the beliefs of each soldier via a method that uses orthogonal summation of probability mass associated with different propositions. Belief propositions are considered exhaustive and exclusive tolerating no logical 'and' [3]. The belief state for soldiers 1-4 at the problem outset is shown in table 1. Note that soldiers 1 and 2 experience more uncertainty than soldiers 3 and 4. Eleven propositions exist for the 4 directional problem encompassing not only the initial four directions, but propositions based on logical 'or' as well as the state of ignorance parameterizing uncertainty. E.g. The problem proposition span contains probabilities associated with direction 1 *or* direction 2 (p1 U p2) as well as uncertainty (theta) as shown in table 1.

	p1	p2	p3	p4	p1Up2	p1Up3	p1Up4	p2Up3	p2Up4	p3Up4	theta
1	0.1	0.3	0.0	0.0	0.2	0.1	0.0	0.1	0.0	0.0	0.2
2	0.3	0.1	0	0	0.1	0.2	0	0.1	0.0	0.0	0.2
3	0.05	0.05	0.2	0.0	0.1	0.1	0.0	0.35	0.0	0.1	0.05
4	0.05	0.05	0.2	0.05	0.2	0.1	0.0	0.1	0.0	0.1	0.15

Table 1: Soldier belief probability mass distributions for the 11 propositions spanning the optimal direction problem. The variable theta signifies uncertainty due to ignorance. The vertical label designates the 4 directions, and the horizontal label designates the 11 propositions.

Information fusion of two non-orthogonal propositions is based on multiplication of probability mass and summation of like probability mass categories. This is followed by multiplicative normalization of the probability mass associated with different propositional categories which emanate from empty set-based probability mass designated by phi [4]. The orthogonal amalgamation template or OAT allows for amalgamation of belief information for soldiers 1 and 2 and then soldiers 3 and 4. This is shown in table 2. Each element of the table delineates the probability mass amalgamation associated with the fusion of the respective proposition mass elements situated along the vertical and horizontal. The virtual soldier beliefs associated with soldiers 1 and 2 and soldiers 3 and 4 are then amalgamated into a group virtual soldier for all soldiers representing the squadron group mind. This helps guide group decision making with respect to which direction to proceed. Based on the belief data shown in table 1, it is not clear which direction is optimal suggesting a rigorous need for information fusion using the OAT.

	M(p1)	M(p2)	M(p3)	M(p4)	M(p1Up2)	M(p1Up3)	M(p1Up4)	M(p2Up3)	M(p2Up4)	M(p3Up4)	M(Theta)
M(p1)	M(p1)	phi	phi	phi	M(p1)	M(p1)	M(p1)	phi	phi	phi	M(p1)
M(p2)	phi	M(p2)	phi	phi	M(p2)	phi	phi	M(p2)	M(p2)	phi	M(p2)
M(p3)	phi	phi	M(p3)	phi	phi	M(p3)	phi	M(p3)	phi	M(p3)	M(p3)
M(p4)	phi	phi	phi	M(p4)	phi	phi	M(p4)	phi	M(p4)	M(p4)	M(p4)
M(p1Up2)	M(p1)	M(p2)	phi	phi	M(p1Up2)	M(p1)	M(p1)	M(p2)	M(p2)	phi	M(p1Up2)
M(p1Up3)	M(p1)	phi	M(p3)	phi	M(p1)	M(p1Up3)	M(p1)	M(p3)	phi	M(p3)	M(p1Up3)
M(p1Up4	M(p1)	phi	phi	M(p4)	M(p1)	M(p1)	M(p1Up4)	phi	M(p4)	M(p4)	M(p1Up4)
M(p2Up3)	phi	M(p2)	M(p3)	phi	M(p2)	M(p3)	phi	M(p2Up3)	M(p2)	M(p3)	M(p2Up3)
M(p2Up4)	phi	M(p2)	phi	M(p4)	M(p2)	phi	M(p4)	M(p2)	M(p2Up4)	M(p4)	M(p2Up4)
M(p3Up4)	phi	phi	M(p3)	M(p4)	phi	M(p3)	M(p4)	M(p3)	M(p4)	M(p3Up4)	M(p3Up4)
M(Theta)	M(p1)	M(p2)	M(p3)	M(p4)	M(p1Up2)	M(p1Up3)	M(p1Up4)	M(p2Up3)	M(p2Up4)	M(p3Up4)	M(Theta)

Table 2: Orthogonal amalgamation template (OAT) for the 11 propositional states spanning the optimal four direction problem. Basic belief assignment is distributed to different propositions at each amalgamation stage using this template. The variable phi signifies the empty set created by the orthogonal summation process and is probability mass ultimately redistributed to all non-empty propositions by a multiplicative normalization factor. The variable phi represents uncertainty based on ignorance or not knowing. The vertical and horizontal labels designate the 11 probability mass propositions existing in the belief span of the problem.

3. Problem Scenario and Methodology for Soldier Squadron Building Clearing and Advancement

In the next problem scenario, a squadron is tasked with moving through a 4-floor building and 'clearing' it. Rooms on a floor are connected sequentially to each other where each floor is connected to the adjacent floor above as shown conceptually in Fig. 1.

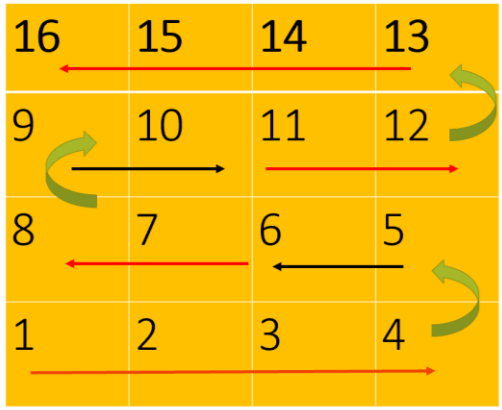

Figure 1: Military floor plan scenario for a squadron ‘clearing’ a two-dimensional building. Height and width dimensions shown. Dynamic programming is used to estimate the optimal points in space when hostile engagements should be performed along each floor of the building. A 16-room building is traversed in the direction outlined where the Bellman equation is optimized over the 2 actions of hostile or non-hostile engagement. A possible policy transition output result is shown in red and black. Red arrows signify hostile transitions and black arrows signify non-hostile transitions. Green arrows are transition points to advance to the next floor of the building.

Dynamic programming is exploited to assist leadership in gaining a broad sense of how simple state policies can be extracted from limited amounts of prior building geo-intelligence. A 16 state-2 action dynamic programming model for building ‘clearing’ is created where the aim is to understand on *average* what optimal points in space should be places of hostile engagement (action 1) and what points should not (action 2) in the 16-room state space domain (Fig. 1). Policy prescription of action based on state emanates from the application of Bellman’s equation [5] which allows for estimation of the present value V of a state s in terms of rewards R, transition probabilities P, and future values V' which are dependent on future states s'. Prior ground clandestine intelligence provides the room-to-room transition probabilistic fields as well as the reward fields. The higher and more important geo-intelligence dynamic programming issue of interest is *how* the optimal policy (function mapping of state to action) changes as *noise* infiltrates the geo-intelligence system, changing the transition probability and reward fields. Understanding this issue is tantamount to understanding *when* estimated policies should be changed based on evidence which modulate the fields.

Baseline transition probability and reward fields are shown in Fig. 2a-d for the 16 state - 2 action domain. State transitions are only possible along the horizontal floors of the building designated as horizontal rows in Fig. 1. Room I signifies the room transitioned to from room J. Transitions to the next floor are only possible at the end points where the green arrows are shown in Fig 1. This transition is quantified in Fig. 2a using large values at the endpoint rooms designated by 4, 8, and 12. When non-hostile engagements are met, room transitions are allowed only along that floor (Fig. 2a) with increases in transition to the end room of the floor as forward progress is made along the floor. Probability for advancement is greatest for rooms which are adjacent. When hostile engagements are met, sufficient defence force and fire power is assumed to allow for defence force victory. This allows for transitions *directly* to the end point of the floor (which is the beginning of the next floor) *irrespective* of room position along the floor (Fig. 2b).

Reward fields for non-hostile engagement and hostile engagement mimic the respective state transition probability field (Fig. 2c-d). When non-hostile engagements are met, rewards are maximum for transitions to the next room in the floor sequence. Smaller rewards are given to the rooms further down in the sequence and to back sliding. Rewards for movement to the first room of the next floor are provided at the end room of the preceding floor. When hostile engagements occur, rewards are given only for transitioning to the end room of the floor and the beginning first room of the next floor (with a slightly lower probability).

a) b)

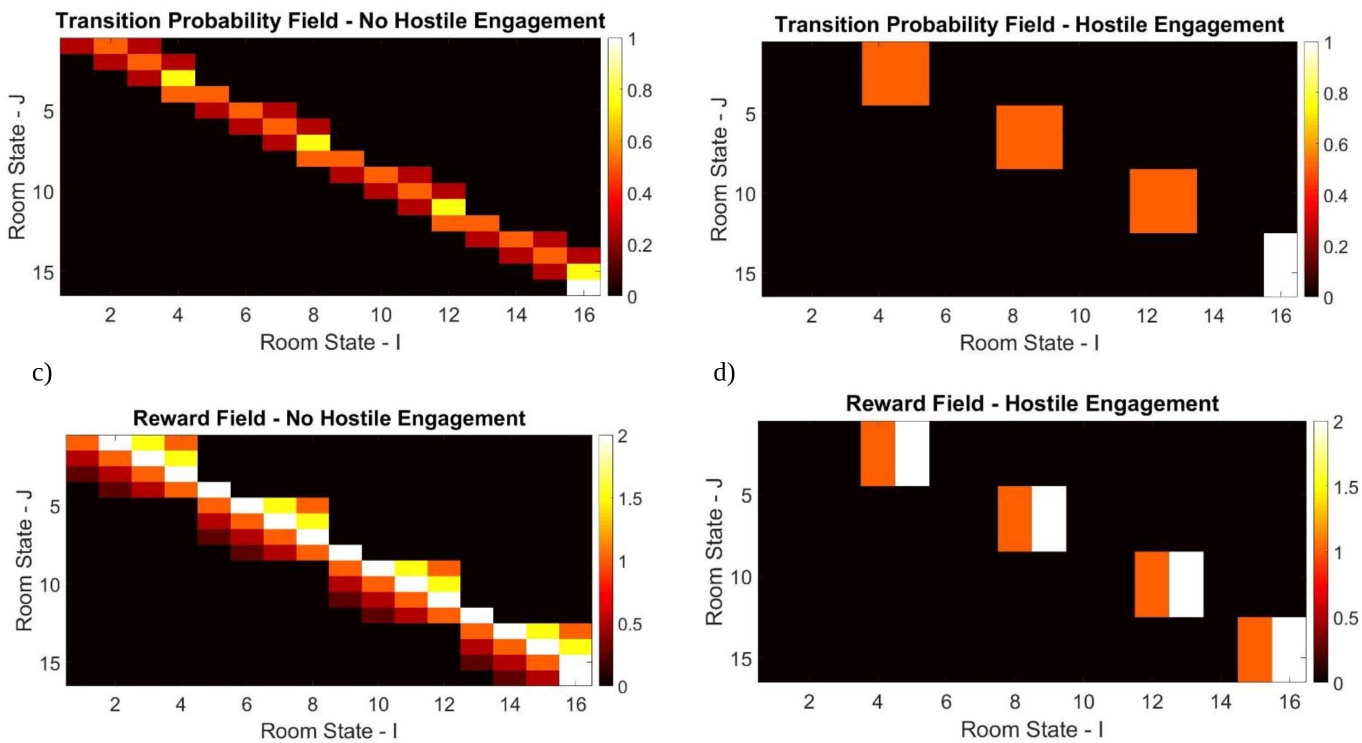

Figure 2: a-b) Transition probability fields for non-hostile engagement and hostile engagements respectively. c-d) Reward field for hostile and non-hostile engagements respectively. Fields span the domain of a 4 floor, 16 room building where the 2 actions of non-hostile and hostile engagements are possible. Note local room transition probability structure and reward fields. Room I signifies the room transitioned to from room J. Low probability is signified by the white colored boxes.

4. Dempster-Shafer Evidential Theory Results for Soldier Squadron Belief Amalgamation

Fig. 3 a-b) shows the probability mass distributions over the 11 propositions for the amalgamations pertaining to soldiers 1-2 and soldiers 3-4 respectively. The virtual mind for soldiers 1-2 shows a maximum at p1 suggesting that direction 1 is optimal for these 2 soldiers. The soldiers 3-4 information amalgamation shows that direction 2 is optimal. Amalgamation for all soldiers shown in Fig. 3c shows that proposition 2 is optimal with proposition 3 being a close second. Proposition 3 being probabilistically a strong option occurs because low but significant probability mass exists for p1 ∪ p2 and p1 ∪ p3 which 'bleeds' probability mass into propositions p2 and p3. Simulation results support the classification result of direction 2 as the optimal direction to proceed in. This result provides geo-intelligence leadership a clear way of discerning belief in a group sense providing a rigorous basis for making a single robust decision.

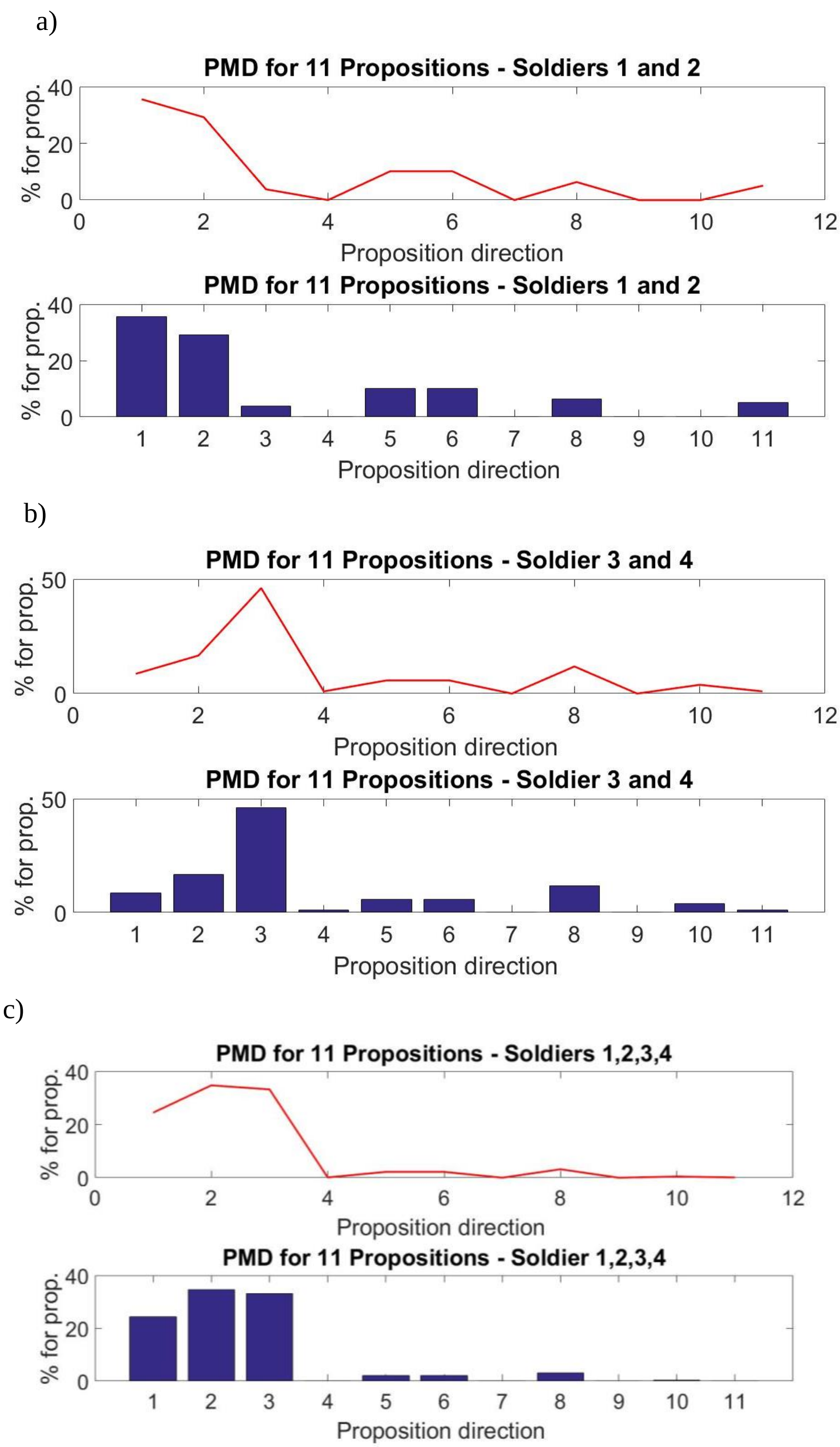

Figure 3: a) Probability mass distribution for soldiers 1 and 2. b) Probability mass distribution for soldiers 3 and 4. c) Probability mass distribution for soldiers 1, 2, 3, and 4. Line plots and bar graphs shown together for clarity. The optimal direction of squadron movement is proposition or direction 2 indicated by the local maximum.

5. Dynamical Programming Results for Soldier Squadron Building Clearing and Advancement Results

The optimal policy resulting from the transition probability and reward fields shown in Fig. 2a-d is displayed in Fig. 4a. It clearly shows that hostile engagement is optimal *along* a floor while passivity is optimal at the end floor transition points labelled as 4, 8, and 12. In other words, there is much to be gained by the squadron through use of judicious hostile engagement where fighting should not be executed in all rooms. This is consistent with floor transitions being choke points where quick movement rather than fighting is rewarded more. E.g. Fighting is easier and more expedient along floors rather than at floor transition points.

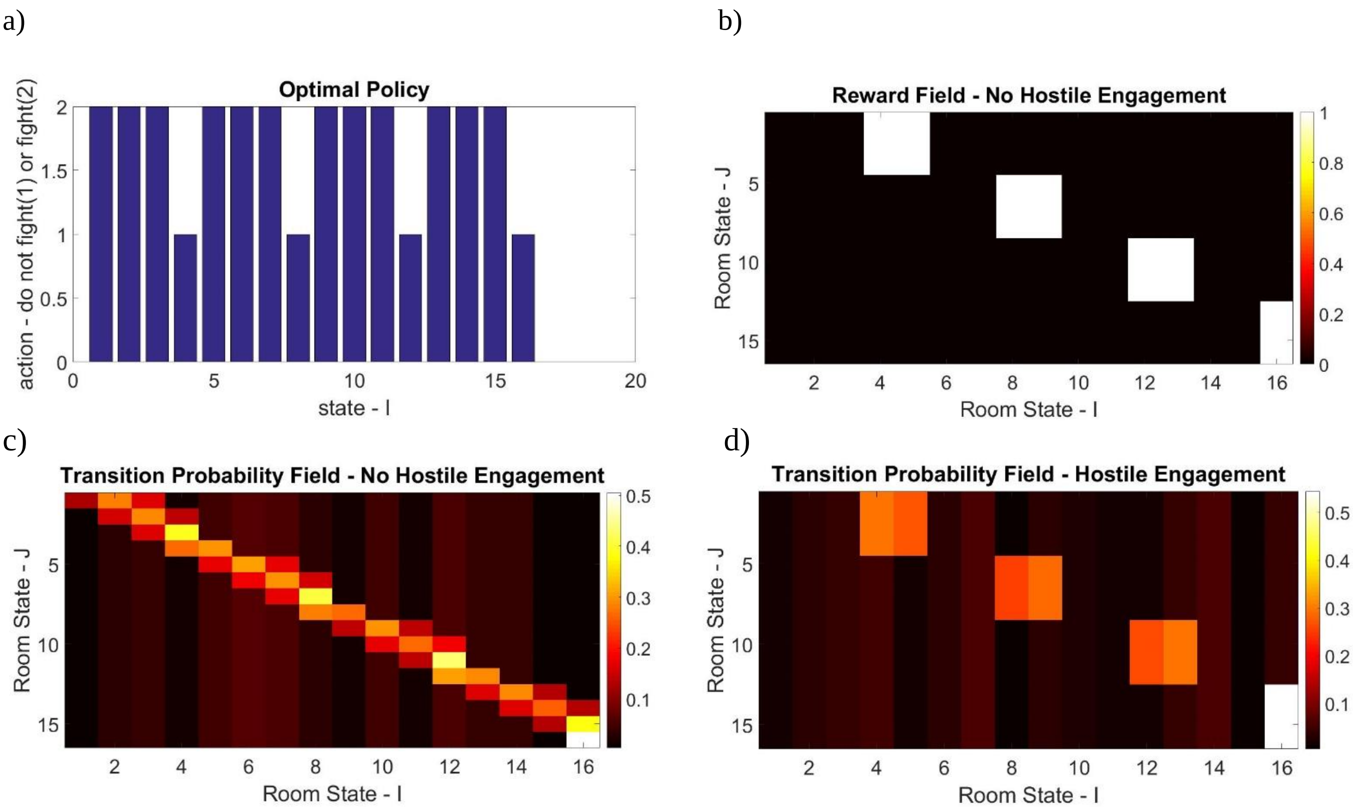

Figure 4: a) Optimal policy for simulation 1 which has Fig. 2a-d) transition probability and reward fields. b) Simulation 2 reward field for non-hostile engagement. Transition probability fields for both actions and the reward field for hostile engagement are the same as simulation 1. c) Simulation 3 transition probability field for non-hostile engagement possessing white noise on top of Fig. 2a non-hostile engagement transition probability field. Mean value and variance for superimposed white noise are 0.0625 and 0.0016 respectively. Transition probability field for hostile engagement is the same as simulation 1 as well as reward fields for both actions. d) Simulation 4 transition probability field for hostile engagement possessing white noise on top of Fig. 2b hostile engagement transition probability field. Mean value and variance for superimposed white noise are 0.0625 and 0.0011 respectively. Transition probability field for no-hostile engagement is the same as simulation 1 as well as reward fields for both actions.

When all fields stay the same and the reward field for non-hostile engagement is lowered near end room floor transition points as shown in Fig. 2b, continual hostile engagement along a floor is no longer optimal. Hostile engagements should be performed near the beginning and middle of the floors as shown in Fig. 5a. This optimal planning simulation suggests that soldier gun artillery should be saved and used only at specific places of hostile engagement along the floor.

a)

b)

c)

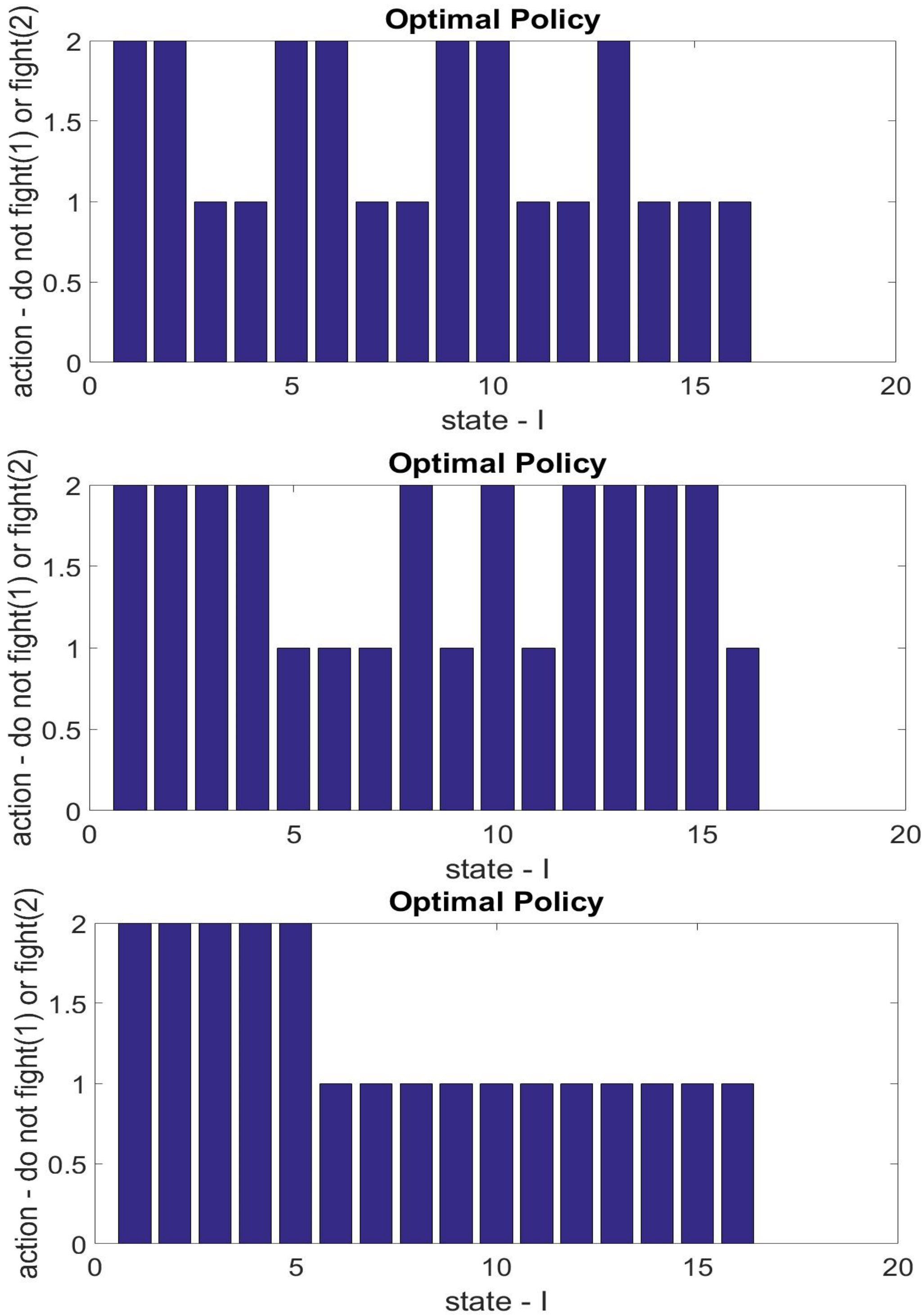

Figure 5: a-c) Optimal policies for simulations 2–4 for the modulated transition probability and reward fields. Actions 1 and 2 delineate no fighting (non-hostile engagement) and fighting (hostile engagement) respectively. See text for discussion.

White noise infiltration of prior geo-intelligence causes modulation of the state transition probability fields which cause changes in the baseline optimal policy shown in simulation 1. White noise was added to the original non-hostile engagement transition probability field in simulation 1 where the mean value and variance were respectively 0.0625 and 0.0016. All the other fields were kept the same as in simulation 1. When white noise is superimposed on the non-hostile engagement transition probability field shown (Fig. 4c), more rooms predicating hostile engagement appear semi-randomly in the building domain (Fig. 5b).

White noise was also added to the original hostile engagement transition probability field in simulation 1 where the mean value and variance were respectively 0.0625 and 0.0011. All the other fields were kept the same as in simulation 1. When white noise is superimposed on the hostile engagement transition probability field as shown in (Fig. 4d), more rooms predicating non-hostile engagement appear semi-randomly in the building domain (Fig. 5c). Both results suggests that uncertainty in the transition probability field associated with one type of action precipitates increases in the alternate action. In other words, increases in hostile engagement actions are associated with increased uncertainty in transition probabilistic information for the non-hostile engagement action. On the other hand, increases in non-hostile engagement actions are associated with increased uncertainty in transition probabilistic information for the hostile-engagement action.

6. Conclusion

D-S evidential theory for optimal direction discernment is demonstrated for a fictious but serious geo-intelligence problem. The formulism can be applied to the development of intelligent squad weapon decision-making aids [6]–[8] to support a *unified* perspective for group decision making. It is especially relevant to problems where ignorance is endemic and where belief states for each soldier are not drastically different. Such differences can create paradoxes which are not only counter intuitive but dangerous where human life is concerned [3].

A simple two-action dynamic programming model is applied to the problem of optimal decision making for 'clearing' a building of hostile adversaries providing leadership rudimentary insight into the probability of fighting engagement along floors. The model provides insight into the effect of uncertainty on policy modulation, allowing for potential understanding as to when and where resources (artillery, increased manpower, and more reconnaissance) are needed. This is especially important when prior situational knowledge is suspect. Expansion of the dynamic programming results is possible via the use of function approximations [9] which would allow for a fuller understanding of action-based policy based on more data.

References

[1] G. Shafer, *A Mathematical Theory of Evidence*. New Jersey, USA: Princeton University Press, 1976.

[2] R. Bellman, *Dynamic Programming*. New York, NY: Dover, 1957.

[3] L. A. Klein, *Sensor and Data Fusion: A Tool for Information Assessment and Decision Making*. SPIE Press, 2004.

[4] K. Sentz and S. Ferson, "Combination of Evidence in Dempster-Shafer Theory," SAND2002-0835, 800792, Apr. 2002. doi: 10.2172/800792.

[5] R. Bellman, "A Markovian Decision Process," *J. Math. Mech.*, vol. 6, no. 5, pp. 679–684, 1957.

[6] R. L. Bobb, J. A. Coady, V. O. Barnard, M. A. Mueller, W. D. Casebeer, and J. P. Salisbury, "Virtual reality framework for design and evaluation of multispectral computer vision algorithms and augmented reality interfaces for enhancing situational awareness in degraded visual environments," in *Virtual, Augmented, and Mixed Reality (XR) Technology for Multi-Domain Operations III*, 2022, p. 13.

[7] J. P. Salisbury, R. L. Bobb, V. O. Barnard, W. D. Casebeer, and D. M. Huberdeau, "Virtual reality framework for multi-human multi-agent adaptive teamwork," in *Proceedings of I/ITSEC 2022*, Orlando, FL, 2022.

[8] B. J. Lance, G. B. Larkin, J. O. Touryan, J. T. Rexwinkle, S. M. Gustein, S. M. Gordon, O. Toulson, J. Choi, A. Mahdi, C. P. Hung, and V. J. Lawheim, "Minimizing data requirements for soldier-interactive AI/ML applications through opportunistic sensing" in *Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications II*, Apr. 2020, vol. 11413, pp. 19-27, doi: 10.1117/12.2564515.

[9] W. B. Powell, *Approximate Dynamic Programming: Solving the Curses of Dimensionality.* Hoboken, NJ: John Wiley and Sons, 2007.