Machine learning aids targeted guidance to trainee's decision-making performance

LTC Peter Nesbitt, PhD & Quinn Kennedy, PhD
Naval Postgraduate School
Monterey, CA
peter.nesbitt@nps.edu, mqkenned@nps.edu

ABSTRACT

Background: Identifying the conditions, doctrinal concepts and specific trainee actions underlying decision performance assists military trainers in applying limited intervention resources. We explore whether a machine learning approach to a human decision learning process can provide targeted intervention guidance. Consequential Learning Assessment (CLA) measures the trainees' ability to sense the state of the environment and take actions that minimize tactical risk without explicit instruction.
Method: We applied the CLA to a computer-based platoon formation decision making task (PFDT), which includes 32 scenarios randomly presented four times (n = 128 trials). For each scenario, there was an effective, acceptable, or poor decision response. This study was approved by the NPS IRB. Thirty participants (11 female) with no prior experience in leading troops in dismounted infantry operations completed the PFDT. We modeled participants as decision agents in a sequence of decisions in which previous decision(s) should inform the current decision. We then examined whether CLA results varied by scenario factor (time of day, terrain height, terrain vegetation, enemy direction, and enemy likelihood) as a measure of learning, defined as the percent of trials on which effective exploration or effective exploitation occurred.
Results: Application of CLA to the PFDT data revealed over 68% of all decisions resulted in effective behavior across scenario factors and supports learning-focused agent assessment. The presence of low light conditions or enemy direction to the front each resulted in 10% more effective decisions across all participants. This type of learning trend across the cohort offers insight on delivery success of factor specific training.
Conclusion: Preliminary results suggest that application of CLA can identify poor decision learners and guide targeted remediation even while the task is in progress. Advances in the computational approach to learning whereby an agent tries to minimize risk when interacting with a complex, uncertain environment can offer insight to human learning systems.

ABOUT THE AUTHORS

Dr. Peter Nesbitt is an Assistant Professor of Operations Research and Associate Dean of the School of Graduate School of Operational and Information Sciences at the Naval Postgraduate School. Much of his research focuses on mathematical programming specializing in network design, resilience and interdiction. Dr. Nesbitt earned PhD in Operations Research from the Colorado School of Mines, a MS in Operations Research from the Naval Postgraduate School and an undergraduate degree in Systems Engineering from the U.S. Military Academy at West Point.

Dr. Quinn Kennedy is a Research Associate Professor in the Operations Research Department and Director of the Healthcare Modeling and Simulation Education Programs at the Naval Postgraduate School. Her work in behavioral science research focuses on optimizing human performance and decision-making and testing the effectiveness of new technologies on human performance and training. Dr. Kennedy earned both her PhD in Psychology and postdoctoral training from Stanford University.

Machine learning aids targeted guidance to trainee's decision-making performance

LTC Peter Nesbitt, PhD & Quinn Kennedy, PhD
Naval Postgraduate School
Monterey, CA
peter.nesbitt@nps.edu, mqkenned@nps.edu

BACKGROUND

Critical to enhancing U.S. military operational and combat effectiveness is improving leader development and decision making. Key to this goal is to effectively train less experienced military decision makers and provide targeted intervention when needed (Cojocar, 2012; Kennedy, Carlson, & Sciarini, 2018; Landsberg et al, 2012; Lopez, 2011; Nesbitt, Kennedy, Alt, & Fricker, 2015; Spain, Priest, & Murphy, 2012; United States Army Combined Arms Center, 2014;).

Military training requires an assessment of progress to guide informed intervention by decision and in terms of factors inherent in a decision scenario that can impact the decision. Identifying the conditions, doctrinal concepts and specific trainee actions associated with poor performance assists trainers in applying limited intervention resources. One way to conceptualize this assessment is as system within which a trainee interacts with an environment in order to achieve a particular goal.

Modeling a trainee interacting with an environment to achieve a goal as a detailed system may require measuring the: 1) sensory perception of their environment, 2) cognitive assessment of available actions by how they affect the future states of the environment, and 3) development of a policy in seeking a goal. Collecting any one of these measures is a challenging task with techniques informed by active research and potentially requiring access to detailed information on the trainee and their history that is not available. There is a need for an assessment method requiring only information available to the trainer. We propose the consequential learning assessment (CLA), which is informed by the field of reinforcement learning. We apply the CLA to a platoon formation decision task in which military participants had to decide which formation should be used based on specific environmental factors (Kennedy, Haley, Niehaus, & Fitzpatrick, 2019).

The general form of this assessment problem exists in the development of artificial decision systems that adapt to their environment. These systems also require a means to frame information collection for the evaluation and selection of available options - essentially an artificial learning system that desires something. The computational approach inherent to *reinforcement learning* provides modeling techniques that explicitly separate the collection of information, its evaluation and estimation of options, and the systematic negotiation between exploration verses exploitation for a decision agent. Biologically inspired, reinforcement learning considers the behavior of a goal-directed agent interacting and learning in an uncertain environment in order to maximize long term benefit. Insight from the field of reinforcement learning in solving the general problem of learning from interaction to achieve goals can inform how we approach and measure human decision-making.

The origins of reinforcement learning are rooting in Psychology. Thorndike (1898) was an early adopter of the term "reinforcement" in his translation of Pavlov's monograph on conditioned reflexes when referring to a strengthening of a pattern of behavior due to an animal receiving a stimulus and culminating in his early Law of Effect, the reinforcement principle according to which rewarded actions tend to be repeated. Turing's (1948) work describing machinery that demonstrates intelligent behavior leverages this Law of Effect as a foundation of artificial intelligence. Further development of the machine learning and specifically reinforcement learning continues to leverage the idea of evaluation of actions taken.

The framework and terminology of reinforcement learning follows (Sutton & Barto 2018). A *reward* is a measure of desirability for the agent, and *regret* is specifically a measure on non-desirability. A *state* is the current conditions in

an environment on which an action is based. An *action* is the decision of an agent with the intent it will gain a reward or change future conditions for improved reward gain. A *policy* is a mapping from perceived state of environment to action to take, similar to stimulus-response rules in psychology. A *value function* generally does not change and defines the long-term intent of the agent. The *model* represents the environment in which the agent perceives and takes action, see Figure 1.

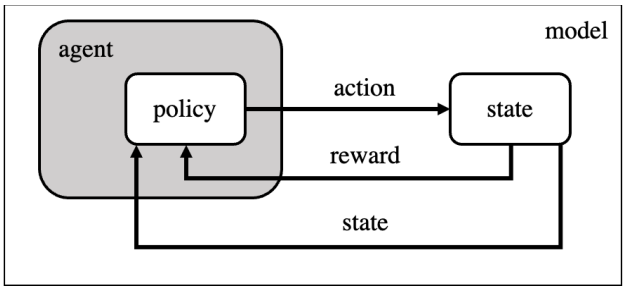


Figure 1. Illustrative example: Reinforcement Learning terminology.

A human-centered example of reinforcement learning is a customer testing a new flavor of ice cream before finalizing the purchase. The customer is the agent that seeks a reward, ice cream worth the purchase. A policy is to see what flavors are available in the shop (the model) and ask for taste tests of unknown flavors before committing to single flavor for purchase, according to the reward function of flavor rating. The value function is satisfaction of the purchase. A similar and example from artificial intelligence is an adaptive controller adjusting thermal parameters of a home (fan, window coverings, furnace, etc.) when evaluated with weather conditions to maintain a comfort zone and minimize energy use. We propose a framework that considers the student as an agent and the training environment as the model.

Consequential Learning Assessment (CLA)

CLA is a means to account for information gained on each presentation by the subject in order to categorize learning success after each decision. CLA assumes the trainee (1) has no knowledge of the specific reward system within the task, (2) seeks to generally apply a value function to sustain an effective path and maximize effective decisions by the end of the task, and (3) has perfect memory. Each decision is evidence of the trainee's level of mastery of the task and categorized in one of four states: i) effective exploration, ii) effective exploitation, iii) ineffective exploration, or iv) ineffective exploitation. The categorization process depends on the feedback gained by the subject within the task before the decision. The decision for a scenario presented for the first time will always categorize as *effective exploration*. Subsequent categorization of decisions account for the trainee's use of feedback. Given feedback of a less than best response option for a particular scenario, any decisions repeating that same scenario - response option pair are categorized as *ineffective exploring*. Once a subject receives feedback confirming the best response option for a particular scenario, there are two possible categorization outcomes. The next time that scenario is presented, again selecting the best response option for that scenario, is categorized as *effective exploitation*; any subsequent decisions selecting other response options for that scenario show regression from the effective path and categorized as *ineffective exploitation*.

We illustrate CLA in Table 1. Table 1 shows possible outcomes for one particular scenario that is randomly presented six times during a training session. For this scenario, response option A is effective; response options B and C are non-effective. The first time the trainee is presented with scenario 1, they select response option B. Because this is the first time they are presented with this scenario, their decision is categorized as effective exploration even though they selected a non-effective option. The second time the scenario is presented, the trainee selects C. Here, their decision also is categorized as effective exploration because they demonstrated that they remembered the negative feedback they received on presentation 1 when they selected B and they had not yet tried C. On the third presentation of this scenario, the trainee selects B; this decision is categorized as noneffective exploration as they already had received negative feedback on this response option but they have not yet selected (and received positive feedback from) A, the effective option. On the fourth presentation, they select A and as this is the first time they have selected it and have received positive feedback on their decision, this decision is categorized as effective exploitation. On the fifth presentation, the trainee selects B and the decision is categorized as non-effective exploitation because they now have received feedback for all three options. On the sixth presentation, they select A and this decision is categorized

as effective exploitation as they are demonstrating using the positive feedback that they received the last time they selected A for this particular scenario.

Table 1. Illustrative example: Application of CLA to a scenario presented six times.

Scenario	Presentation	Action	Feedback	Effectiveness assessment	Explore/exploit assessment
1	1	B	negative	effective	explore
1	2	C	negative	effective	explore
1	3	B	negative	non-effective	explore
1	4	A	positive	effective	exploit
1	5	B	negative	non-effective	exploit
1	6	A	positive	effective	exploit

METHODS

We applied the CLA to previously collected data in which subjects completed the platoon formation decision task (PFDT). This study was approved by the Naval Postgraduate School IRB and United States Marine Corps Human Research Protection Program.

Subjects

Thirty subjects (26 military personnel; 11 female) with no prior experience in leading troops in dismounted infantry operations completed the PFDT. Mean age was 38.20 (10.04) years. Military personnel came from all services (2 Army, 6 Marines, 12 Navy, 4 Air Force, 2 Coast Guard) with an average of 13.60 (5.80) years of service, with 22 of the 26 having previously deployed.

Platoon formation task (PFDT)

The PFDT (Kennedy, et al., 2019) was designed and piloted tested by military subject matter experts at the Naval Postgraduate School. It has a 2^5 factorial design, resulting in 32 scenarios. For each scenario, there is an effective, acceptable, or poor decision response. Each scenario is randomly presented four times for a total of 128 trials. The five factors are: time of day (daylight or twilight), terrain height (flat or hilly), terrain vegetation (scrub brush or dense trees), enemy direction (front or side), and enemy likelihood (possible or likely). During each trial, subjects watch a 10 second first person view video moving through a terrain scenario. Three factors are manipulated in the videos (time of day, terrain height, and terrain vegetation). Enemy direction and likelihood information are provided in text format below the video. Three platoon formation options are provided on the right side of the screen (see Figure 2). Subjects make their decision by clicking on one of the platoon formations. Immediately after making their decision, the subject receive one of two messages: “You made the effective choice.” or “You did not make the effective choice.” The message remains on the screen until the subject selects the “NEXT” button and the next scenario begins. Software outputs a file that includes scenario, platoon formation decision, time to make each decision for each trial.

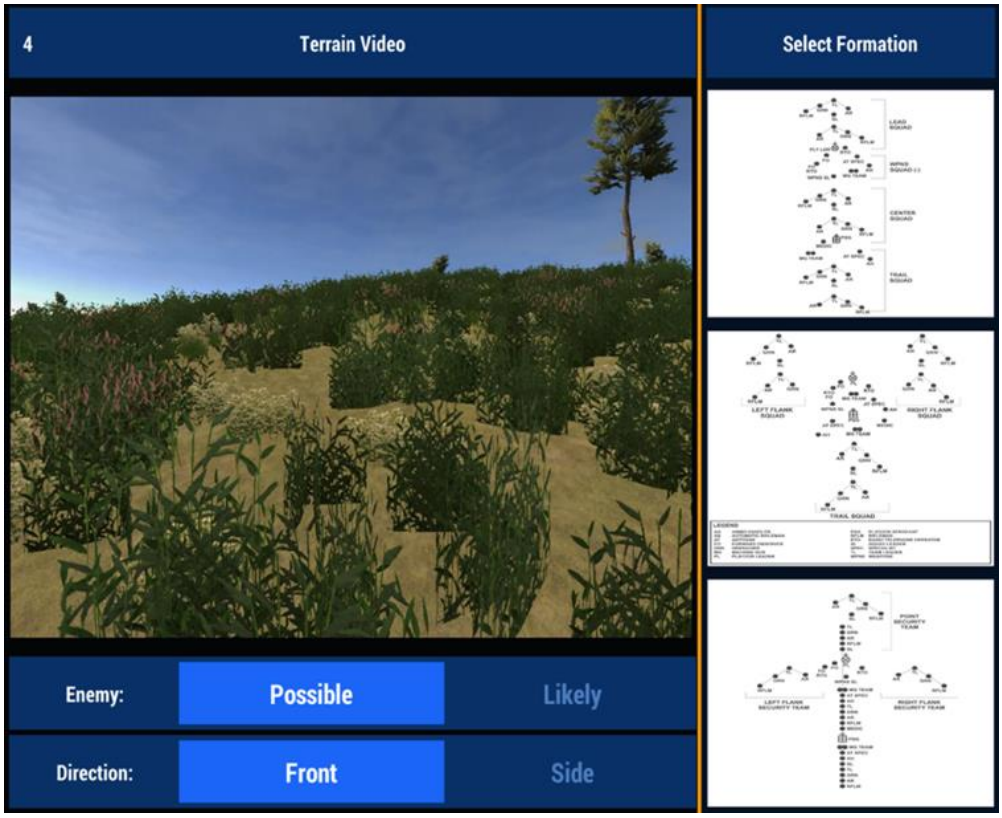


Figure 1. Snapshot of a PFDT scenario.

CLA Application to the PFDT

Using the method described above, we calculated the four categories of CLA, effective exploration, ineffective exploitation, effective exploitation, and ineffective exploitation at three levels of analysis: scenario, tactical action, environmental factor, and individual subject. We define a measure of *overall proficiency* as the ratio in the number of scenarios in which a subject achieves effective exploitation and exploration to the number of scenarios in which they demonstrate ineffective exploitation and exploration. Higher scores indicate greater proficiency.

Software

We employed R version 4.2.0 (2022-04-22) for CLA data configuration.

RESULTS

Application of CLA to the PFDT data revealed over 68% of all decisions resulted in effective behavior and supports learning-focused agent assessment. The presence of low light conditions or enemy direction to the front each resulted in 10% more effective decisions across all participants. This type of learning trend across the cohort offers insight on delivery success of factor specific training. Next, we applied CLA to the PFDT data at three different levels: by overall group proficiency, by an individual factor, and by individual learner.

CLA results at the overall group level There was a wide range of overall proficiency scores (41.7-1.5), see Table 2 on the next page. We then used cluster analyses reveal two groups based on overall proficiency scores.

Table 2. CLA results for PFDT by subject: Overall proficiency cluster group and action results.

	overall	assigned	effective exploitation			effective exploration			ineffective exploitation			ineffective exploration		
subject	proficiency	cluster	column	file	vee	column	file	vee	column	file	vee	column	file	vee
230	41.7	2	57	31	31	6	0	0	1	1	1	0	0	0
280	20.3	2	56	31	25	4	1	5	1	0	1	3	0	1
136	9.7	2	52	22	31	7	3	1	3	3	0	2	4	0
123	9.7	2	52	30	27	2	2	3	10	0	2	0	0	0
200	8.1	2	48	21	31	8	5	1	2	4	0	6	2	0
192	7.5	2	53	22	19	7	4	8	4	2	2	0	4	3
108	7.0	2	43	28	28	7	2	4	3	1	0	11	1	0
250	7.0	2	48	28	26	5	3	2	3	0	4	8	1	0
164	7.0	2	46	31	20	7	1	7	6	0	4	5	0	1
187	6.5	2	40	32	29	9	0	1	5	0	1	10	0	1
220	6.1	2	39	32	26	11	0	2	2	0	1	12	0	3
109	6.1	2	33	29	32	14	2	0	3	1	0	14	0	0
175	5.7	2	45	21	25	8	6	4	5	1	1	6	4	2
176	5.7	2	50	23	26	6	4	0	2	2	6	6	3	0
135	4.8	2	43	31	21	8	1	2	2	0	9	11	0	0
111	4.6	2	47	19	24	7	5	3	3	3	3	7	5	2
163	3.9	2	45	19	24	6	3	5	3	4	1	10	6	2
290	3.9	2	48	26	9	8	4	7	3	0	5	5	2	11
191	3.4	1	47	19	24	6	2	1	1	9	5	10	2	2
188	3.4	1	39	22	17	11	5	5	7	2	9	7	3	1
112	3.1	1	46	19	18	8	3	3	5	2	9	5	8	2
210	3.0	1	35	17	30	11	3	0	12	7	2	6	5	0
148	3.0	1	31	17	23	13	7	5	9	8	4	11	0	0
240	2.8	1	35	7	27	14	7	4	11	9	1	4	9	0
147	2.7	1	40	24	15	5	3	6	16	3	5	3	2	6
124	2.3	1	39	15	20	6	4	5	15	1	3	4	12	4
158	2.2	1	38	18	12	7	6	7	11	7	4	8	1	9
159	2.2	1	31	11	24	14	6	2	10	8	6	9	7	0
270	1.8	1	31	14	12	15	7	4	8	4	13	10	7	3
260	1.5	1	38	9	1	10	9	10	8	4	3	8	10	18

We can identify further grouping in support of training intervention with hierarchical cluster analysis, see Figure 3. Initially, each subject is assigned to its own cluster and then the algorithm proceeds iteratively, at each stage joining the two most similar clusters by subject performance of their final decision by scenario, continuing until there is just a single cluster.

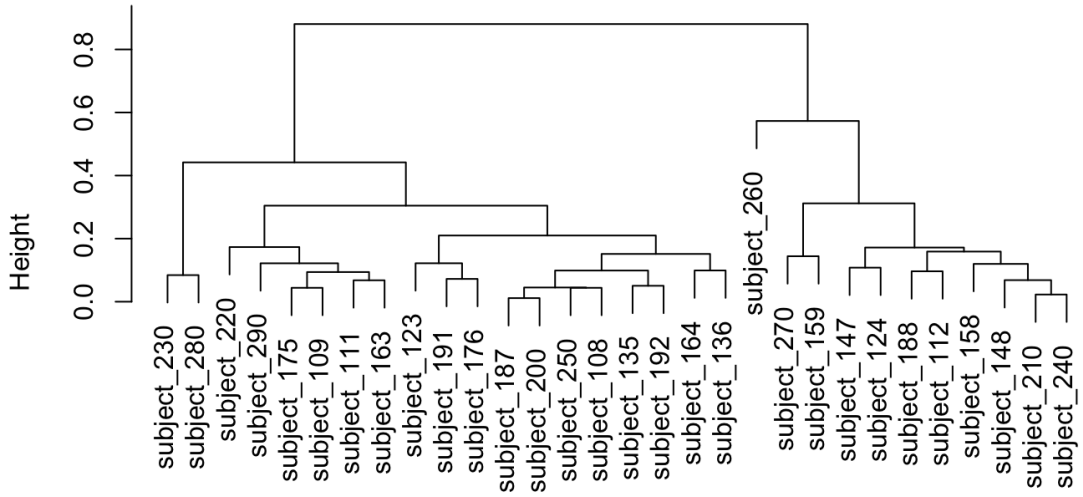


Figure 3. Hierarchical cluster analysis for subgroup analysis: the depicted cluster dendrogram indicates which subjects made similar decisions to each other.

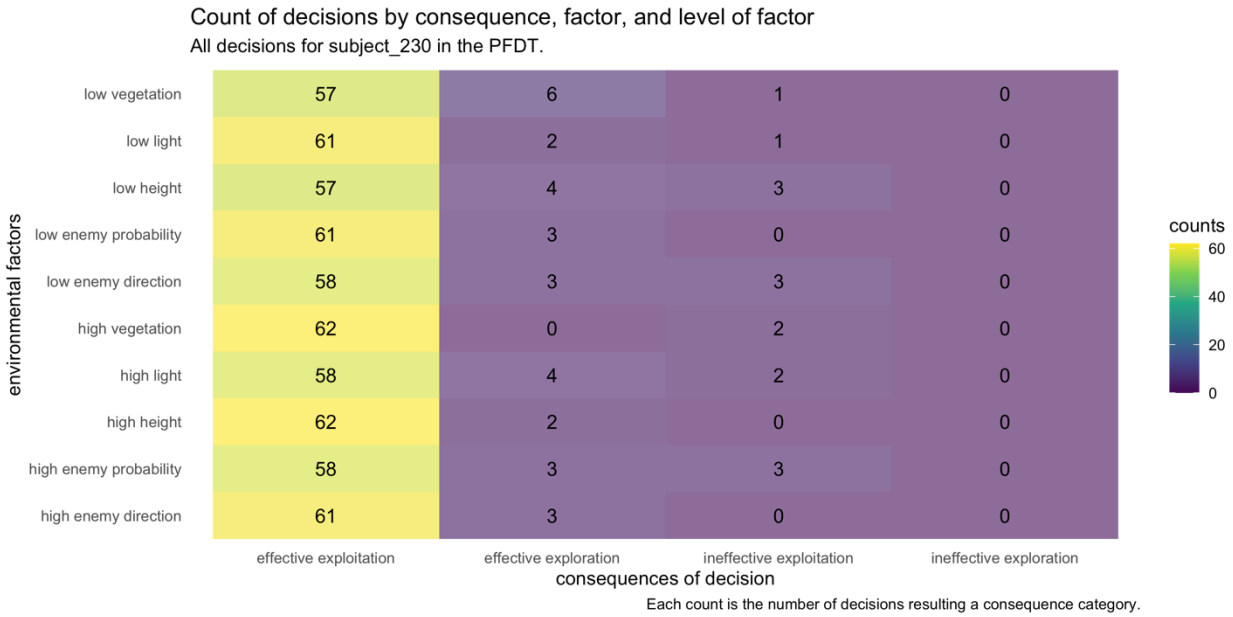
CLA results at the factor level. To demonstrate how CLA can be used at the factor level, we calculated CLA categories for each subject for the 64 scenarios in which the factor level, high vegetation, was depicted. Effective exploitation scores ranged from 22 – 62; ineffective exploitation ranged from 0 – 17. Providing frequencies of effective and ineffective exploration adds additional insights. As shown in Table 3, subject 230 took two decisions under exploration followed by 62 (97%) effective exploitative decisions. In contrast, subject 270 made an equivalent number of effective and ineffective exploration decisions, making effective exploitative decisions on only 54% of the scenarios with high vegetation. We see some subjects perform better by factor: subject 200 is tied for third for scenarios that contain high vegetation, whereas they are fifth rated by overall performance. Similarly, subject 159 improves beyond their overall performance by two ranks for scenarios that contain high vegetation.

Table 3. Count of each subject's CLA performance for all PFDT trials in which the factor of vegetation height was high (n = 64), ordered by descending count of effective exploitation responses.

subject	effective exploitation	effective exploration	ineffective exploitation	ineffective exploration
230	62	2	0	0
280	55	6	0	3
136	53	6	3	2
123	54	3	7	0
200	55	6	2	1
192	43	10	6	5
108	50	7	2	5
250	54	4	1	5
164	48	8	4	4
187	50	4	4	6
220	47	7	1	9
109	45	10	1	8
175	45	7	5	7
176	52	6	3	3
135	47	7	6	4
111	48	7	2	7
163	45	5	6	8
290	40	10	6	8
191	45	5	7	7
188	37	13	9	5
112	43	6	8	7
210	38	8	11	7
148	30	17	8	9
240	37	13	9	5
147	37	9	11	7
124	42	7	7	8
158	35	7	13	9
159	36	8	17	3
270	34	10	11	9
260	22	16	3	23

CLA results by individual subject

Using CLA, a trainer can investigate the effects of specific environmental factors on an individual subject's decision-making. We illustrate this by depicting specific CLA results for two subjects. Figure 4a uses color to depict how the consequences of a subject's decisions may vary for each factor level. We see that no factor or level of factor impacts the effectiveness of subject 230's decisions. In comparison, subject 260 shows a much more varied decision profile, demonstrating a tendency towards ineffective exploitation. Figure 4b magnifies the subtle influences of changes of factor level on an individual's decision. Even though subject 230 performed very well, their decision performance was most impacted by changes of vegetation level. Subject 260 shows consistent decision making across level changes of enemy probability but shows a lack of sensitivity to changes in factor level for enemy direction, vegetation and light.



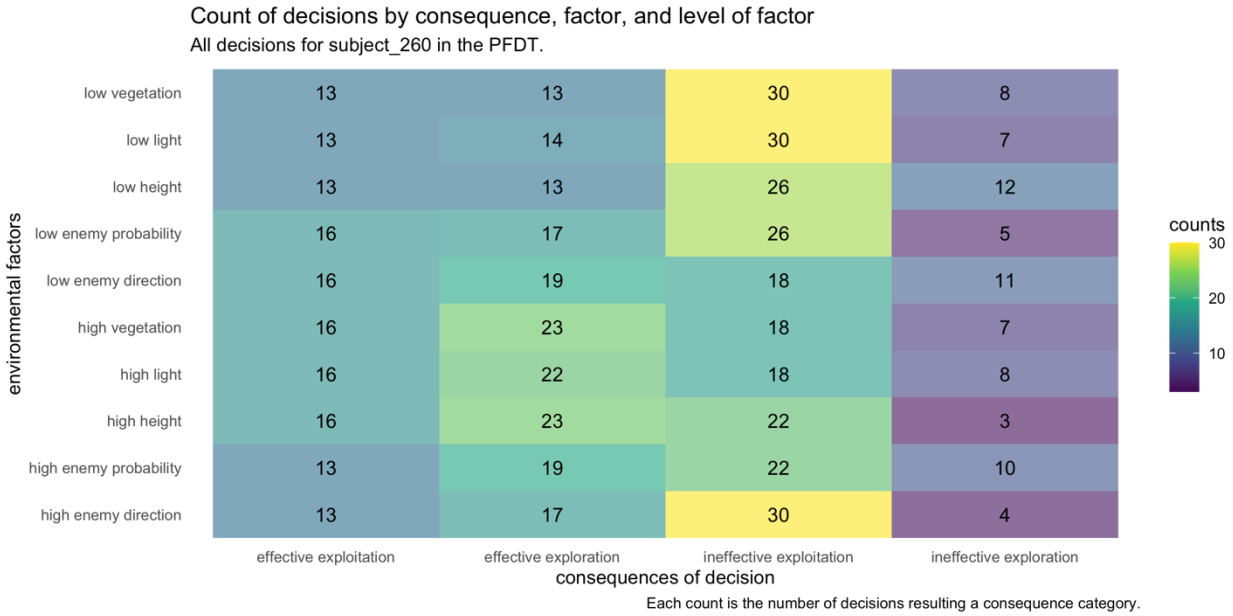
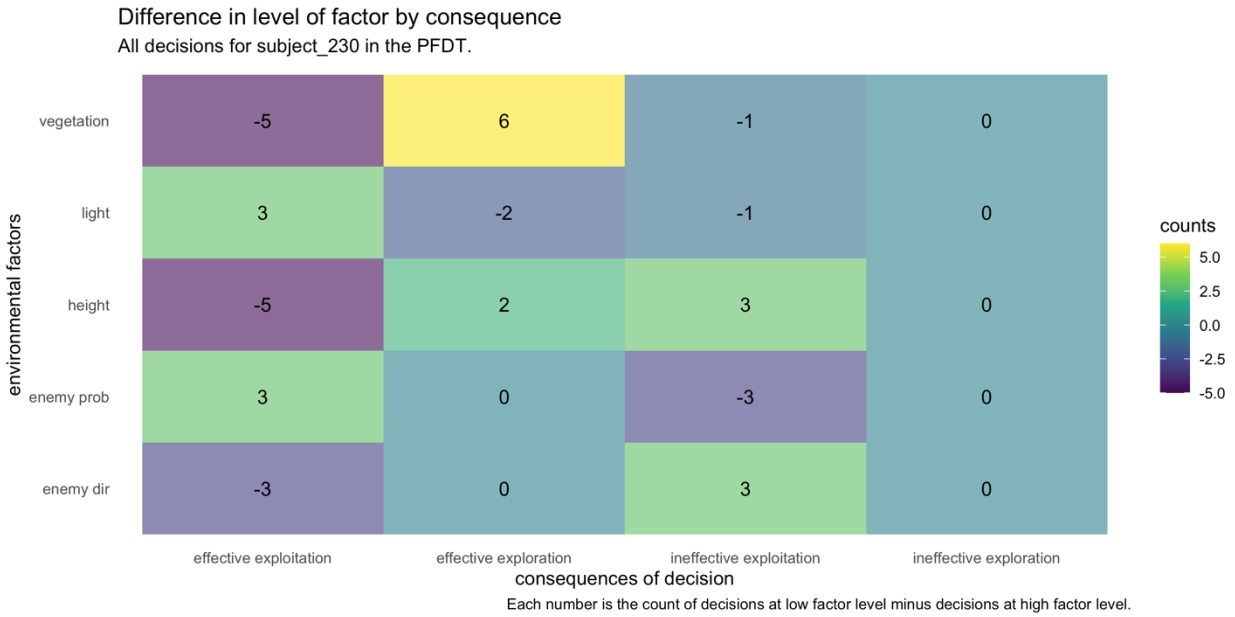


Figure 4a. Counts of decision consequence: a tally of a single subject's decisions by factor level and consequence



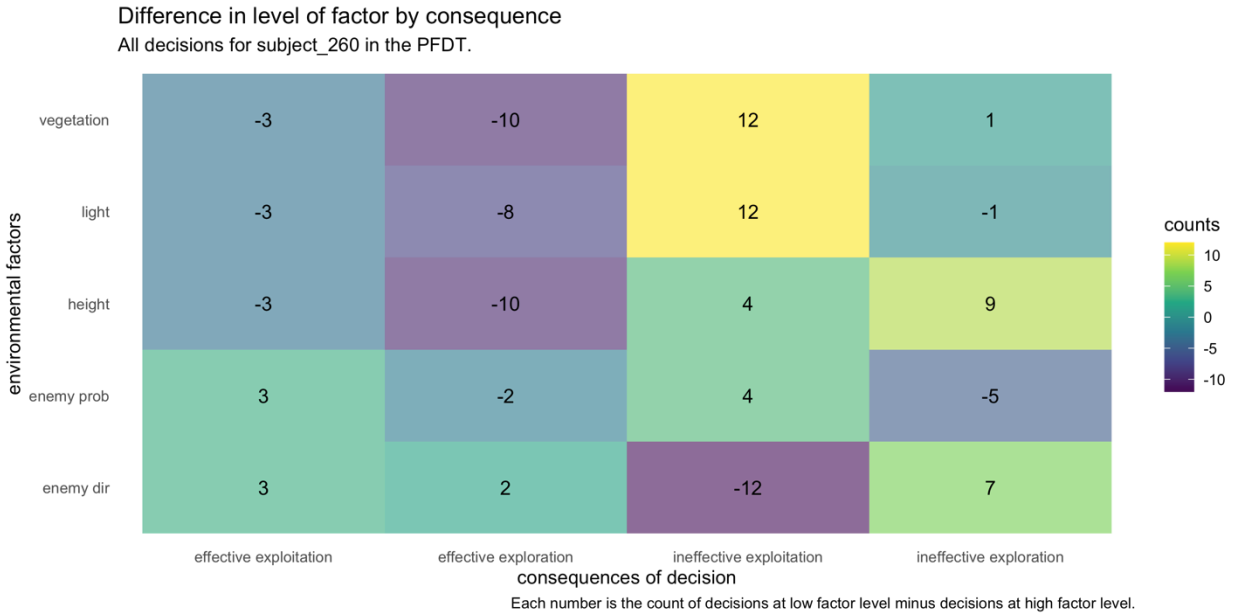


Figure 4b. Intra-level differences of consequence: counts represent the difference in decision frequency within each factor (low factor level – high factor level)

DISCUSSION

Using the principles of reinforcement learning, CLA accounts for an exploration component of the decision process that is excluded from the typical decision performance measure of percent correct. We demonstrate CLA can provide informative results to the trainer at a global, factor and individual subject levels. CLA assumes there are four different consequences to a subject's action: effective exploitation, effective exploration, ineffective exploitation, and ineffective exploration. CLA maintains record of each state-action pair as well its effectiveness in context with learning on earlier actions in the task.

The strengths of this approach are that it incorporates an exploration component of the decision process that is excluded from the typical decision performance measure of percent correct. Application of CLA allows for factor analysis and investigation of critical components of the task and subject performance. Additionally, CLA only requires the typical data output generated from the task and not any additional output. However, we note the necessary complexity of the data structure. This technique is only as good as the decision task for which its being used. It can verify the subject performance according to the task but cannot validate the task itself.

We tested the CLA on the PFDT, but it can be used to measure decision performance for a variety decision training tasks that require multiple trials. CLA may not be a good choice for high impact low frequency events, which do not lend themselves to the application of reinforcement learning. Future study is to test the effectiveness of CLA on a task that includes an adversary.

Military trainers need to know which trainee can stop because they have achieved proficiency, which trainee needs to continue training, and which trainee requires an instructor for intervention. CLA can provide this information to military trainers.

ACKNOWLEDGEMENTS

This work was supported by the Office of Naval Research Code 34.

REFERENCES

Cojocar, W.J. (2011). Adaptive leadership in the military decision making process. *Military Review* (Nov-Dec), 23 - 28.

Kennedy, Q., Carlson, T. & Sciarini, L. (2018) Preliminary validation of an adaptive tactical training model: Cognitive alignment with performance targeted training intervention model (CAPTTIM). In H. Ayaz and F. Dehais (Eds) *Neuroergonomics: The Brain at Work and in Everyday Lif*e (pp 127 – 131). Elsevier: San Diego, CA.

Kennedy, Q., Hanley, B., Niehaus, J., & Fitzpatrick, C. (2019). *Accelerating training of platoon formation decisions through a computer-based task*. American Psychological Association Annual Convention, Chicago, IL.

Landsberg, C., Astwood Jr., R., Van Buskirk, W., Townsend, L., Steinhauser, N., & Mercado, A. (2012) Review of Adaptive Training System Techniques, *Military Psychology*, 24(2), 96-113.

Lopez, T. (2011). Odierno outlines priorities as Army Chief. *Army News Service.*, Retrieved from http://www.defense.gov/News/NewsArticle.aspx?ID=65292s.

Nesbitt, P., Kennedy, Q., Alt, J., & Fricker, R. (2015). Iowa Gambling task modified for military domain. *Military Psychology, 27*(4), 252 - 260.

Spain, R., Priest, H., & Murphy, J. (2012). Current trends in adaptive training with military applications: An Introduction. *Military Psychology*, 24(2), 87-95.

Sutton, R. S., & Barto, A. G. (2018). Reinforcement learning: An introduction. MIT press.

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. The Psychological Review: Monograph Supplements, 2(4).

Turing, A. M. (1948). Intelligent machinery.

United States Army Combined Arms Center (2014) *The Human Dimension White Paper*