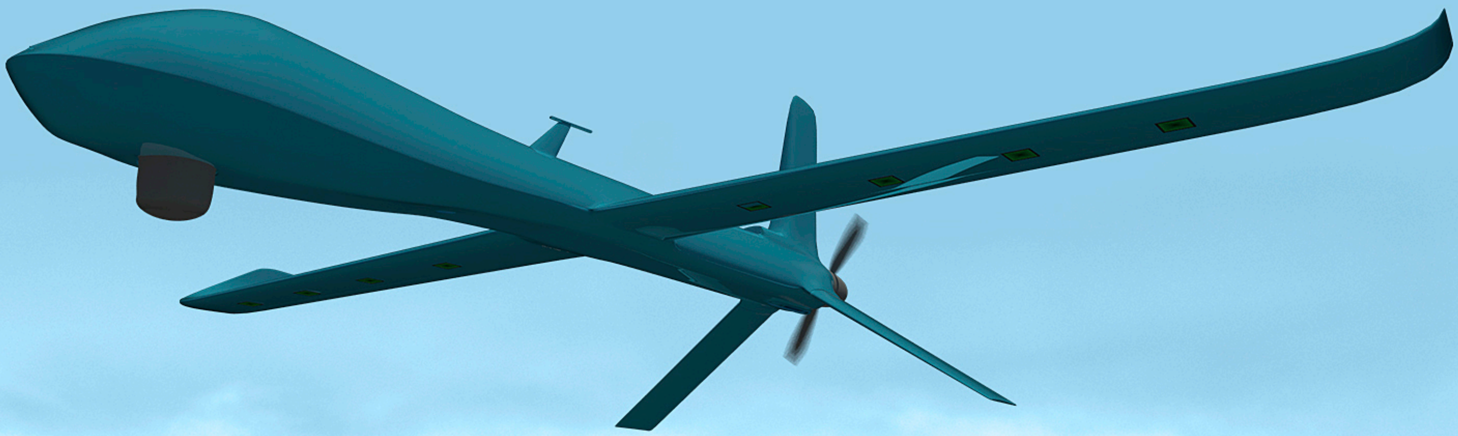
Dedicated to innovation in aerospace

NLR-TP-2017-267 | February 2018

Towards Modeling the Learning Process of Aviators using Deep Reinforcement Learning

CUSTOMER: Ministry of Defence



NLR – Netherlands Aerospace Centre

Netherlands Aerospace Centre

NLR is a leading international research centre for aerospace. Bolstered by its multidisciplinary expertise and unrivalled research facilities, NLR provides innovative and integral solutions for the complex challenges in the aerospace sector.

NLR's activities span the full spectrum of Research Development Test & Evaluation (RDT & E). Given NLR's specialist knowledge and facilities, companies turn to NLR for validation, verification, qualification, simulation and evaluation. NLR thereby bridges the gap between research and practical applications, while working for both government and industry at home and abroad.

NLR stands for practical and innovative solutions, technical expertise and a long-term design vision. This allows NLR's cutting edge technology to find its way into successful aerospace programs of OEMs, including Airbus, Embraer and Pilatus. NLR contributes to (military) programs, such as ESA's IXV re-entry vehicle, the F-35, the Apache helicopter, and European programs, including SESAR and Clean Sky 2.

Founded in 1919, and employing some 650 people, NLR achieved a turnover of 71 million euros in 2016, of which three-quarters derived from contract research, and the remaining from government funds.

For more information visit: **www.nlr.nl**

EXECUTIVE SUMMARY

UNCLASSIFIED

Towards Modeling the Learning Process of Aviators using Deep Reinforcement Learning

Problem area

In this publication we report on our study of the performance of Deep Reinforcement Learning (DRL) agents in performing tasks that are illustrative for human Sensor Operators (SOs) in Remotely Piloted Aircraft Systems (RPASs). Our hypothesis is that the descriptive and predictive qualities of the agent's learning process potentially allow us to identify human task requirements, training needs, selection criteria and cut-off benchmarks.

Description of work

We (1) constructed tasks (games) that require some of the abilities of the SO, (2) devised DRL agents that have to learn these tasks and (3) performed initial learning tests with DRL agents on these tasks. In constructing the tasks, we took a two-pronged approach. First we attempted with a state-of-the-art DRL algorithm to master a complex game, the Space Fortress (SF) game, designed by psychologists and previously used for e.g. training and selection in the aviation domain. Second, we constructed simple tasks (mini-games) that impose less demands on memory

REPORT NUMBER
NLR-TP-2017-267

AUTHOR(S)
J. van Oijen
G. Poppinga
O. Brouwer
A. Aliko
J.J.M. Roessingh

REPORT CLASSIFICATION
UNCLASSIFIED

DATE
February 2018

KNOWLEDGE AREA(S)
Training, Mission Simulation and Operator Performance

DESCRIPTOR(S)
Machine Learning
RPAS
Sensor Operator
Training
Serious Games

capabilities, inferring unknown rules of the game and higher-order dynamical control. Three types of mini-games were constructed (navigation, shooting and tracking) each with a base-line task and three variations of increasing complexity. These address abilities of (1) visual tracking and spatial processing, (2) vigilance to multiple sources and divided attention and (3) discrimination respectively.

Results and conclusions

We present DRL results on tasks that cover different cognitive abilities required for an SO, using games as a method for learning. The full SF game could not be learned by our DRL agent using the A3C-LSTM algorithm. Learning proved to be mainly based on suppressing undesirable behavior, which in turn is the result of negative rewarding. Reward signals are not frequent enough to move the parameters of the DRL agent towards the correct gradient. Results on the SO mini-games showed that the base-line and the variation addressing visual tracking and spatial processing showed super- or par-human performance (within the limited world size used). The two more complex variations addressing divided attention and discrimination could not be mastered (though improved scores have been observed towards the end of the training). In a side experiment, positive transfer of training has been observed through progressive part-task training (up to 20% score increase in equal total training time).

In conclusion, this work is a small step towards testing our initial hypothesis: to what extend can the descriptive and predictive qualities of a DRL agent's learning process be fostered for human factor challenges such as identifying human task requirements, training needs, selection criteria or cut-off benchmarks.

Applicability

This work focuses on using DRL agents to learn complex tasks that are typical for aviation. Identifying similarities between the learning curves of humans and of DRL agents may allow NLR to use DRL agents to make predictions for its customers on the effectiveness and efficiency of selection, training and task environment for human operators.

Follow-on trials will focus on comparison of the properties of human learning curves with those of DRL agents. These properties of the learning curve are, for example, initial performance level, slope of the learning curve and asymptotic performance level. In addition, similarities in transfer-of-training between DRL agents and humans will be investigated.

GENERAL NOTE

This report is based on a paper published in the IEEE Systems, Man and Cybernetics (SMC) conference proceedings, Banff, Canada, 5-8 October 2017.

NLR
Anthony Fokkerweg 2
1059 CM Amsterdam
p) +31 88 511 3113 f) +31 88 511 3210
e) info@nlr.nl i) www.nlr.nl

Dedicated to innovation in aerospace

NLR-TP-2017-267 | February 2018

Towards Modeling the Learning Process of Aviators using Deep Reinforcement Learning

CUSTOMER: Ministry of Defence

AUTHOR(S):

J. van Oijen	NLR
G. Poppinga	NLR
O. Brouwer	NLR
A. Aliko	NLR
J.J.M. Roessingh	NLR

NLR - Netherlands Aerospace Centre

This report is based on a paper published in the IEEE Systems, Man and Cybernetics (SMC) conference proceedings, Banff, Canada, 5-8 October 2017.

The contents of this report may be cited on condition that full credit is given to NLR and the authors. This publication has been refereed by the Advisory Committee AEROSPACE OPERATIONS (AO).

CUSTOMER	**Ministry of Defence (MoD)**
CONTRACT NUMBER	**01614104801 – NLR Program Flying 2020 for MoD**
OWNER	**NLR**
DIVISION NLR	**Aerospace Operations**
DISTRIBUTION	**Unlimited**
CLASSIFICATION OF TITLE	**UNCLASSIFIED**

APPROVED BY :		
AUTHOR	**REVIEWER**	**MANAGING DEPARTMENT**
J. van Oijen	J. van der Pal	H.G.M. Bohnen
DATE 1 5 0 1 1 8	DATE 1 5 0 1 1 8	DATE 2 0 0 1 1 8

Contents

I. INTRODUCTION 5

II. LITERATURE 5

A. *Human Factors in RPAS Operations* 5

B. *Job Requirements for a Sensor Operator* 5

C. *Gaming* 6

D. *Machine Learning* 6

III. CONNECTING ABILITIES, GAMING AND MACHINE LEARNING 6

IV. REINFORCEMENT LEARNING APPLIED TO SO TASKS 7

V. EXPERIMENTS AND RESULTS 8

A. *Space Fortress* 8

B. *SO-mini games* 8

VI. CONCLUSIONS AND FUTURE WORK 9

VII.REFERENCES 10

This page is intentionally left blank.

Towards Modeling the Learning Process of Aviators using Deep Reinforcement Learning

Joost van Oijen, Gerald Poppinga,Olaf Brouwer, Andi Aliko, Jan Joris Roessingh
Air Operations Division
Netherlands Aerospace Center NLR
Amsterdam, the Netherlands

***Abstract*—In this paper we report on our study of the performance of Deep Reinforcement Learning (DRL) agents in performing tasks that are illustrative for human Sensor Operators (SOs) in Remotely Piloted Aircraft Systems (RPASs). Our hypothesis is that the descriptive and predictive qualities of the agent's learning process potentially allow us to identify human task requirements, training needs, selection criteria and cut-off benchmarks. We present DRL results on tasks that cover different cognitive abilities required for an SO, using games as a method for learning.**

I. Introduction

Modelling human learning through neural processing has a long history, tracing back to Hebbian Learning in the forties of last century [1] . However, only recently, Deep Reinforcement Learning (DRL) agents became capable of learning and mastering a range of 'vintage' video games, based on merely a series of video images/frames, as the output of the agent's own actions, which are either reinforced or weakened through reward signals (acquired game points) [2]. This inspired us to explore the potential of DRL networks to tackle "real-life" tasks to investigate the learning process of aviators.

The basic idea of this study is that if DRL agents can be used to construct a representative model for human learning, this model can then be applied for the purpose of personnel training, personnel selection and task design. A practical example would be to predict usability and trainability in case of changes to the working environment. Also, these agents could be deployed to determine optimal transfer-of-training between different training environments, as will be tentatively examined in this study. In terms of selection, DRL agents could be used as a reference for performance on selection tasks. Furthermore, DRL agents can be deployed to support humans in their working environment. To realize such potential applications, still much research is required regarding human learning models, the potential use of machine learning to simulate learning, as well as technological advancements of DRL algorithms. This study is a small step on this road exploring the potential of current DLR algorithms to successfully play serious games designed for human learning.

In this paper we describe how machine learning has been applied to explore learning models of (sub) tasks relevant for Sensor Operators (SOs) in Remotely Piloted Aircraft Systems (RPAS). Serious games are used as a vehicle to assess DRL on its capability of learning certain cognitive abilities that are required for an SO in executing its tasks. In performing this assessment our aim is to gain insight into the potential applicability of current (and future) DRL algorithms in supporting human factor challenges in training and selection.

The paper is structured as follows. Section II provides background literature on job analysis of SOs, gaming and machine learning. A connection between these three fields is proposed in section III. In section IV we describe our approach of applying DRL on illustrative SO tasks using games. Section V outlines performed experiments and results after which we conclude in section VI.

II. Literature

A. Human Factors in RPAS Operations

Remotely Piloted Aircraft System (RPAS) have become increasingly valuable military assets, and reliance upon RPAS operations will continue to increase [3]. Operating RPAS systems require crews to fulfill new roles and tasks [4]. For example, the pilot remotely operates and navigates the aircraft during the en-route, mission, and return phase of flight. The sensor operator (SO) operates the on-board electro optical and infrared sensors from the ground control station.

A variety of human factor challenges has been associated with RPAS operations and human operator roles. For instance, both pilots and SOs are prone to experience sensory isolation and reduced attention. Addressing such challenges is crucial to improve job performance and safety. Solutions include the design of suitable selection- and training methods and an improved working environment.

B. Job Requirements for a Sensor Operator

Two major approaches exist in job analysis, being task-oriented and worker-oriented [5]. The task-oriented approach focuses on the actual activities and tasks involved in performing the work [6]. Examples of SO tasks include monitoring sensors, selecting and providing data to analysts, detecting, identifying and tracking targets [7] [8] [9]. The worker-oriented approach focuses on abilities needed to perform a job successfully [6] [10]. Examples of abilities associated with SOs include multitasking, spatial processing, and memory [4]. These abilities refer to intellectual mental functions and information processing abilities essential to SO job accomplishment [11]. 0 provides an overview of six core cognitive abilities related to SO job performance. These include cognitive proficiency, visual perception, attention, spatial processing, memory, and reasoning [4] [11]. SOs who fall short on these abilities are prone to experience degrade job performance. Reference [12] found that working memory and logical reasoning are relatively strong predictors with respect to training progression and task performance of SOs.

TABLE1 COGNITIVE ABILITIES OF A SO

Cognitive Proficiency	General cognitive ability
	Speed & accuracy of information processing
Visual Perception	Visual acuity, scanning & discrimination
	Visual recognition, tracking & analysis
Attention	Vigilance to multiple sources of visual & auditory information (situational awareness)
	Sustained & divided attention to visual & auditory information
Spatial Processing	Spatial analysis & orientation
	Spatial reasoning & construction (manipulation of 2-dimensional information into 4-dimensional mental imagery
Memory	Visual & auditory memory (working, immediate & delayed)
	Spatial memory (working, short-term & delayed)
Reasoning	"Real-time" general and deductive reasoning (problem solving)
	Quickly assess risk, likely outcomes & potential repercussions (forward thinking)
	Quickly perceives the next steps and multitasks high level of information & procedures (task prioritization & management)

The abilities illustrated above are referenced in the remainder of this paper when relating cognitive abilities to gaming and deep reinforcement learning.

C. Gaming

The use of games in selection and training of aviators has been studied for decades. When considering the domain of RPASs, the control environment bears large similarity with a video game environment. Gaming experience has been found to transfer the performance of RPAS pilots with respect to flight tasks [13] [14].

A prime example is the game of *Space Fortress* (SF) [15]. This arcade-style game from the 80's was developed by psychologists to study complex skill acquisition and has shown positive transfer of training to actual flight performance [16]. A screenshot of SF is illustrated in Fig. 1. The goal of the game is to destroy a fortress while evading and destroying mines. An identify friend or foe (IFF) procedure has to be performed in order to be able to destroy the foe mines: based on alphanumeric characters that were briefly displayed at the start of the game, the player has to switch to the appropriate type of weapon for foe mine destruction. The game addresses a complex task that is representative of real-life tasks and is sufficiently difficult and challenging to keep the task interesting for human subjects during extended practice [15]. It can take up to 20 hours of training for humans to reach expert level. Not only is the game demanding in terms of perceptual, cognitive and motor skills required, it also requires knowledge of the rules and game strategy.

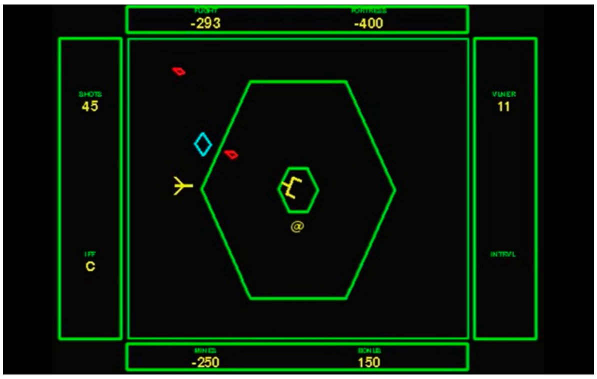


Fig. 1. Screenshot of Space Fortress

D. Machine Learning

Machine learning is a subfield of artificial intelligence that focuses on providing computers with the ability to learn a task without being explicitly programmed. Learning takes place supervised, unsupervised or reinforced through interactions with an environment [17]. In this work we focus on the last method which is known as reinforcement learning. More specifically we consider deep reinforcement learning (DRL) applied to game environments, a field which has shown considerable progress in recent years. In [2], Mnih et al. introduced a Deep Q-Network (DQN) that allowed DRL agents to play a diverse collection of Atari 2600 Games with, sometimes, super human performance. Agents merely used the screen pixels as input and learned how to play the game by performing actions and adapting the network based on the scoring progress of the game. This work has sparked much follow-up research addressing shortcomings such as dealing with exploration issues [18], sparse rewards [19] or partial observability [20], hereby increasing the performance of DRL agents in a wider variety of games and tasks.

III. CONNECTING ABILITIES, GAMING AND MACHINE LEARNING

In this study we analyze the performance of DRL agents in mastering certain cognitive abilities through gaming. This requires two aspects to be addressed: (1) the design of a suitable game that addresses the respective cognitive abilities and (2) the extent to which DRL would be able to learn those abilities through gaming.

Addressing the former, instead of designing a new game, we limit ourselves to the *Space Fortress* (SF) game introduced in the previous section. SF is treated as a suitable candidate for this study since it has shown positive transfer of training [16] and has been linked to many of the cognitive abilities required for an SO as described in Table I. For instance: visual acuity, scanning & discrimination [21], vigilance to multiple sources of visual & auditory information [22] or spatial memory [23].

Addressing the latter, an assessment has been made on the characteristics of DRL algorithms, focusing on their capability of demonstrating behavior that could be associated with the cognitive abilities required for an SO while playing computer games. 0displays a summary of this assessment.

TABLE II ASSESSMENT OF DRL CAPABILITIES.

Cognitive Proficiency	For highly limited contexts with very specific tasks, DRL suggests a general cognitive ability (such as when playing Atari 2600 games). In relation to Atari style games, DRL is only applied for processing visual information. DRL seems capable of rapid and accurate visual information processing and in specific cases displaying super human performance, as demonstrated in gameplay on various Atari games. In some other instances, the performance of DRL is clearly below human performance.
Visual Perception	Visual acuity can be related to a neuroscientific basis for convolutional networks, as indicated in [24]. Where the human eye has a limited area of the field of view in high resolution, the network convolutes the complete screen in one single resolution. Discrimination is done implicitly and likely to depend on the extent to which a visual object is influential for maximizing the score. DRL can learn to recognize objects that are relevant for selecting the actions that maximize the score. Tracking does not take place explicitly, but DRL may be able to "predict" movement through Recurrent Neural Network (RNN) components (e.g. LSTM), and adapt its behavior accordingly.
Attention	In successfully learned games, DRL seems capable of handling multiple sources requiring attention at the same time, and behaves so to maximize a reward function. Combining multiple sources also seems possible; an example of a network simultaneously comparing multiple images as input is a Siamese network [25]. In successful DRL gameplay, the network does seem to display sustained and divided attention, as e.g. illustrated by t-SNE "saliency maps" in e.g. [2] . However, the relevance and hierarchy of the different types of attention can be very difficult to learn. This is an area of research on structuring of attention, e.g. with a hierarchy of tasks in feudal networks [26].
Spatial Processing	The initial layers of the visual oriented DRL networks convolve the visual input in time, and thus relate to spatial analysis, orientation and movement and map spatial characteristics to actions. Seemingly dependent on the complexity of the physics, DRL can learn how to play the game. If the physics are more complex (e.g. acceleration based movement), the learning appears to become significantly more difficult. If the environment is not fully observable, spatial reasoning is expected to be unlikely as there is no built in model in the networks to reason with.
Memory	The recurrent neural layer of a network could be regarded as immediate visual memory. The transitions stored in RNN components could be considered as the spatial context of the agent, but there is no explicit model. Delayed memory is not present, and working memory is an area of research [27].
Reasoning	It seems possible to successfully combine deep neural networks with general and deductive reasoning techniques (e.g. tree search), to make use of Monte-Carlo simulations and historic game play, as has been displayed with AlphaGo [28]. The problem of task prioritization and management exploration is central in reinforcement learning. There is much research in this field, e.g. in Hierarchical DQN [19] and feudal networks [26]

Although the above assessment suggests that DRL networks would be capable to display certain abilities, still it is seen that its performance can highly depend on the complexity of the tasks, the action-space of the agent and the state-space of the environment within the game. Considering the game of SF, preliminary analysis suggests it may be too challenging to reach human-level performance. For instance, when considering the DQN algorithm described in [2], it was reported that the game *Asteroids,* a game which has similar game play elements to SF, only achieves 7% of human-level performance. However, several improved algorithms have been proposed in recent times that may achieve better results.

IV. REINFORCEMENT LEARNING APPLIED TO SO TASKS

In the previous section, a qualitative assessment was performed on the capabilities of DRL with respect to the cognitive abilities of an SO. To investigate these capabilities in a game setting, experiments are conducted following two approaches. In the first approach we attempt to apply DRL to the full game of SF using state-of-the-art algorithms. This approach can be seen as whole-task learning in which DRL is confronted with learning multiple abilities at once required to play the game Experimental results of this approach are described in the next section. In the second approach we break down the problem domain into so-called mini-games to investigate the performance of DRL. The mini-games are designed to address one or more selected cognitive abilities that would be required to master such games. Further, game elements are incorporated that are illustrative for sub-tasks of an SO. At a later stage, separate tasks can be combined to represent more complex cognitive tasks. This cumulative approach towards learning tasks is also seen in humans and is known as part-task training. For instance, in [29], the SF game was decomposed into separate sub-tasks, trained human subjects on the tasks and then verified the performance of the subjects on the overall game. Experimental results indicated that training on the sub-games made the subjects perform better on the overall game.

Fig. 2 provides an overview of the mini-games that have been proposed in order to investigate the performance of DRL on illustrative SO tasks and associated cognitive abilities. They can be categorized across two dimensions, namely *task-oriented* and *worker-oriented*.

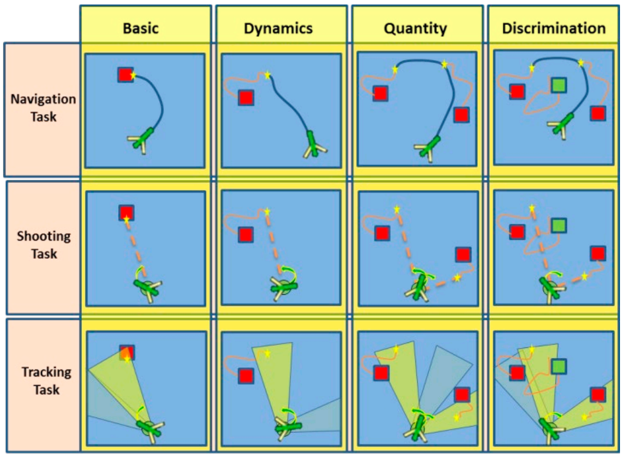


Fig. 2. Overview of the Sensor Operator mini-games

1) *Task-oriented:* In Fig. 2, different rows represent different task-oriented game elements, consisting of a navigation task, a shooting task and a tracking task. In the navigation task the player's goal is to intercept and avoid enemies and friendlies respectively. In the shooting task the goal is to shoot enemies (and prevent friendly fire). In the tracking task the goal is to track targets with a sensor beam. These tasks can roughly be related to different tasks of an SO: operating the camera to navigate and orient; to aim and operate weapon systems; and to acquire targets, track targets and perform FFI-identification (friend or foe).

2) *Worker-oriented:* In Fig. 2, different columns represent worker-oriented game elements. These are variations to the task-oriented games and increase or decrease the cognitive proficiency that is required. The variations are (1) *dynamics* (differentiating between static and moving targets); (2) *quantity* (varying the number of targets); and (3) *discrimination* (differentiating between enemies and friendlies). These variations can be related to the cognitive abilities that have been identified in 0. For instance, the inclusion of moving targets affects visual perception (visual acuity) and spatial processing; increasing the amount of targets affects attention (vigilance to multiple sources and divided attention); and differentiating between enemies and friendlies affects visual discrimination. Further, all three variations relate to cognitive abilities such as speed and accuracy of information processing. Note that each variation builds upon the previous variation, thus adapting task complexity in an incremental manner. The specific ordering of the variations that was chosen could also have been different.

V. EXPERIMENTS AND RESULTS

In this section we describe the experiments that have been performed and their results on applying DRL on the SF game (top-down approach) and SO-mini games (bottom-up approach) respectively.

A. Space Fortress

The game of SF was introduced in section II.C. This arcade style game displays similarities with the Atari 2600 games that were used in DRL research [2]. Applying DRL to the SF game, initially the open-source library of *SimpleDQN* was employed [30]. This library was developed with the goal to replicate DeepMind's results from [2]. An *OpenAI Gym* environment [31] was created to connect the Deep Q-learning Agent and the game framework. Subsequently *GoogleDQN* [32] was used for learning to play SF. The networks for both DQN methods did not learn to play the game and did not show significant progress during training. As a consequence, sub-tasks of the game were identified and trained separately, namely an aiming task and a navigation task. Learning progress was seen but still clearly far below human performance.

As a next effort a more recent DRL algorithm was tried, namely A3C+LSTM [18].This algorithm is said to improve in situations where extensive spatio-temporal planning is required (as is the case in SF). After training for an extensive period, the results were very limited. Within SF, the tasks that achieve positive rewards are extremely sparse. An example of the complexity of getting rewards lies in the unique method for destroying the fortress: only when firing ten times with an interval of at least 250 msec. and then a consecutive 11th time within 250 msec. leads to significant reward. Note that for humans playing the game, such a game rule is known beforehand.

Analysis on the results suggest that the 'punishments' players receive for undesired behavior limits the network's exploration in behavior that could lead to the (sparse) rewards. However, through negative reinforcement alone, the agent was still able to learn a few things. The agent was able to quickly suppress the firing behavior: when not suppressed, it resulted in many consecutive negative rewards when firing while empty. It also learned a moving pattern that is slightly better than random behavior in terms of total reward per game.

In order to investigate learned flying behavior compared to random play, we removed any action output that includes firing from the random play test and observed the rewards. The flying behavior, as manifested by the total rewards per game, still seems to be noticeably better on average. An average score of -2350 was observed throughout learning, whereas random play, and random play with firing disabled achieved -11106 and -3026 respectively. Successful exploration of a hierarchically complex reward space is a field of very active research. We believe that games such as SF will greatly benefit from this research in the coming years.

B. Sensor Operator-mini games

All twelve mini-games from Fig.2. have been implemented in a custom game framework. In all tasks (navigation, shooting and tracking), the *basic* variation includes one static enemy which becomes dynamic in the *dynamics* variation. Dynamic objects follow a predictable trajectory and do not exhibit AI themselves. In the *quantity* variation, two dynamic enemies are used whereas in the *discrimination* variation, a dynamic friendly is introduced. When the player reaches a goal (intercepts, shoots or tracks), the corresponding target is respawned in the world at a new random location. To give insight into the state space of these games, the world size is 7x7 units; the object size is one unit with a 0.1 units per step moving speed. Player rotation is done in 11.25 degrees per step. This relatively small state space is used since early experiments have shown that large state spaces require significantly longer training time for DRL algorithms.

Similar to the top-down effort of SF, *SimpleDQN* was employed [30]. Agent observations were encoded as game frames of 48x48 pixels in gray-scale. Rewards for the agent during training directly relate to an increase or decrease of the game's score (e.g. enemy hit results in +1, friendly hit results in -1). Training on each game was performed for at least 25 million training steps (frames) which corresponded to about 30 hours of training (faster than real-time) per mini-game on the computers used.

0illustrates the results of the experiments for each task (navigation, shooting and tracking) and for each variation. Training results are compared to human scores based on the average of five iterations of playing a game (consisting of 5000 steps). Agent scores are expressed as percentages compared to human scores. Training results are used from the best performing epoch (learning iteration). As a baseline, scorings of an agent performing random actions are included.

TABLE III SO MINI-GAMES RESULTS

Navigation Task				
	Basic	*Dynamics*	*Quantity*	*Discrimination*
Agent score	**109%**	**91%**	**66%**	**22%**
Random score	3%	18%	30%	13%
Shooting Task				
	Basic	*Dynamics*	*Quantity*	*Discrimination*
Agent score	**120%**	**101%**	**76%**	**71%**
Random score	6%	21%	25%	16%
Tracking Task				
	Basic	*Dynamics*	*Quantity*	*Discrimination*
Agent score	**102%**	**102%**	**96%**	**92%**
Random score	18%	15%	21%	13%

Looking at the results, in the *basic* and *dynamics* variations the agent reaches super-human or (near) par-human performance in all tasks. The only significant super-human performance is seen in the basic shooting task in which the agent excels in reaction time and accuracy. In the *quantity* and *discrimination* variations, the agent performs sub-human and performance starts to degrade in relation to the complexity of the task. However, agent performance still increases at the end of the training time which suggests opportunities for an agent to improve when trained longer.

A difference in difficulty in learning a task is also clearly observed from the agent's performance for each task. In the navigation task, the lowest performance is seen, followed by the shooting task, followed by the tracking task. We believe this difference is due to the difficulty for an agent to obtain a reward for a task. In the navigation task, random actions in the early training phases rarely lead to rewards. In the tracking task, rewards are immediate and much easier to obtain.

In an attempt to decrease the training time for complex mini-games an additional experiment was performed. The goal of this experiment was to verify if training time of a complex variation could be reduced through progressive part-task training. Training results from a less complex variation were used as input for a more complex variation. Research has shown that such a strategy is beneficial for human learning [29]. We considered the navigation/discrimination task. Initial (non-part-task) training on this task reached 54% human performance after 75 million frames. Alternatively, this task was trained incrementally by first training the *dynamics* variation, followed by the *quantity* variation and concluding with the *discrimination* variation. Training time of part-tasks was divided equally (hereby reaching the same amount of training time). Results showed an increase of 20% of the score, resulting in 65% human performance. This suggests positive transfer of training, such that less overall training time is required when tasks are learned incrementally. Fig. 3. shows the learning progress of the *quantity* and *discrimination* variations with and without pre-training. The influence the specific ordering of variations has with respect to incremental part-task training has not been investigated and is left for future work.

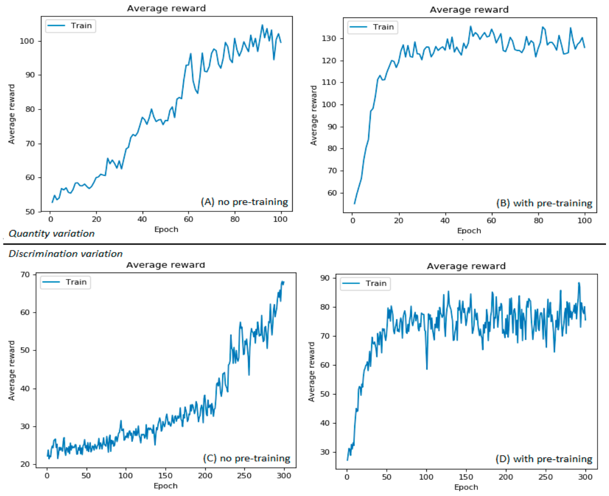


Fig. 3. Progressive part-task training results for the navigation task: comparing no pre-training (A,C) versus pre-training (B,D)

VI. CONCLUSIONS AND FUTURE WORK

This work is a small step towards testing our initial hypothesis: to what extend can the descriptive and predictive qualities of a DRL agent's learning process be fostered for human factor challenges such as identifying human task requirements, training needs, selection criteria or cut-off benchmarks. As a first step, in this study we explored the potential of DRL agents in performing tasks which are illustrative for RPAS sensor operators. Gaming was used as a connection: on the one hand, games can be used by humans to enhance cognitive abilities that can be transferred to real-life tasks. On the other hand, recent developments in DRL algorithms have shown promising results in their application to games. We (1) constructed tasks (games) that require some of the abilities of the SO, (2) devised DRL agents that have to learn these tasks and (3) performed initial learning tests with DRL agents on these tasks.

In constructing the tasks, we took a two-pronged approach. First we attempted with a state-of-the-art DRL algorithm to master a complex game, the *Space Fortress* (SF) game, designed by psychologists and previously used for e.g. training and selection in the aviation domain. Second, we constructed simple tasks (mini-games) that impose less demands on memory capabilities, inferring unknown rules of the game and higher-order dynamical control. Three types of mini-games were constructed (navigation, shooting and tracking) each with a base-line task and three variations of increasing complexity. These address abilities of (1) visual tracking and spatial processing, (2) vigilance to multiple sources and divided attention and (3) discrimination respectively.

The full SF game could not be learned by our DRL agent using the A3C-LSTM algorithm. Learning proved to be mainly based on suppressing undesirable behavior, which in turn is the result of negative rewarding. Reward signals are not frequent enough to move the parameters of the DRL agent towards the correct gradient. Results on the SO mini-games showed that the base-line and the variation addressing visual tracking and spatial processing showed super- or par-human performance (within the limited world size used). The two more complex variations addressing divided attention and discrimination could not be mastered (though improved scores have been observed towards the end of the training). In a side experiment, positive transfer of training has been observed through progressive part-task training (up to 20% score increase in equal total training time).

In conclusion, our DRL agents have been able to successfully learn abilities in different mini-games of very limited complexity. Higher efficiency in learning was seen when using progressive part-task training. In more complex task settings such as the Space Fortress game, current DRL algorithms fall short. In further research we plan to compare the properties of human learning curves (e.g. starting level, slope and asymptote) as well as the conditions for transfer learning (transfer-of-training) with DRL agents and humans.

VII. References

[1] D. Hebb, The organization of behavior: A neuropsychological approach, John Wiley & Sons, 1949.

[2] Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., et al., "Human-level control through deep reinforcement learning," *Nature,* vol. 518, pp. 529-533, 2015.

[3] Murray, C.C., Park, W., "Incorporating Human Factor Considerations in Unmanned Aerial Vehicle Routing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 43, no. 4, pp. 860 - 874, 2013.

[4] Howse, R.W., "Knowledge, Skills, Abilities, and Other Characteristics for Remotely Piloted Aircraft Pilots and Operators," Air Force Center for Applied Personnel Studies (AFCAPS), Randolph AFB TX, USA, 2011.

[5] M. Brannick, E. Levine and F. Morgeson, Job and work analysis: Methods, research, and applications for human resource management, Thousand Oaks, CA: Sage, 2007.

[6] M. Robinson, "What is job analysis," Institute of Work Psychology, 2001.

[7] Gugerty, L., DeBoom, D., Walker R., Burns, J., "Developing a simulated uninhabited aerial vehicle (UAV) task based on cognitive task analysis: task analysis results and preliminary simulator performance data," in *Proceedings of the human factors and ergonomics society 43rd annual meeting*, 1999.

[8] Cohen, M.S. , Freeman, J.T., Wolf, S., "Metarecognition in Time-Stressed Decision Making: Recognizing, Critiquing, and Correcting," *Human Factors,* vol. 38, no. 2, pp. 206 - 219, 2016.

[9] USAF, "Remotely Piloted Aircraft Sensor Operator (RPA SO) - Career Field Education And Training Plan (CFETP)," Department of The Air Force, Washington, 2009.

[10] P. Muchinsky, Psychology applied to work: an introduction to industrial and organizational psychology, Summerfield, NC: Hypergraphic Press, 2012.

[11] Chappelle, W., McDonald, K., King, R.E., "Psychological Attributes Critical to the Performance of MQ-1 Predator and MQ-9 Reaper U.S. Air Force Sensor Operators," Air Force Research Laboratory, Brooks City-Base, TX, 2010.

[12] Roos, C., Boland, E.J., Roessingh, J.J.M., "UAS Ground Control Station manning study: Task analysis, analysis of competencies and experiment," Netherlands Aerospace Center NLR, Amsterdam, the Netherlands, 2010.

[13] Triplett, Captain J., "The effects of commercial video game playing: a comparison of skills and abilities for the predator UAV - thesis," Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, 2008.

[14] McKinley, R.A., McIntire, L.K., Funke, M.A., "Operator Selection for Unmanned Aerial Systems: Comparing Video Game Players and Pilots," *Aviation, Space, and Environmental Medicine,* vol. 82, no. 6, pp. 635 - 642, 2011.

[15] Mane, A., Donchin, E., "The Space Fortress Game," *Acta Psychologica, 71,* pp. 17-22, 1989.

[16] Gopher, D., Weil, M., Bareket, T. , "Transfer of skill from a computer game trainer to flight," *The Journal of the Human Factors and Ergonomics Society,* vol. 36, no. 3, pp. 387-405, 1994.

[17] C. Bishop, Pattern Recognition and Machine Learning, New York: Springer, 2007.

[18] Mnih, V., Puigdomenech Badia, A., Mirza, M., Graves, A., Lillicrap, T.P., Harley, T. et al., "Asynchronous methods for deep reinforcement learning," in *International Conference on Machine Learning*, New York, 2016.

[19] Kulkarni, T.D., Narasimhan, K.R., Saeedi, A., Tenenbaum, J.B., "Hierarchical Deep Reinforcement Learning: Integrating Temporal Abstraction and Intrinsic Motivation," 2016. [Online]. Available: https://arxiv.org/pdf/1604.06057.pdf. [Accessed 10 03 2017].

[20] M. Hausknecht and P. Stone, "Deep Recurrent Q-Learning for Partially Observable MDPs," arXiv preprint, 2015.

[21] Blumen, H.M., Gopher, D., Steinerman, J.R., Stern, Y., "Training Cognitive Control in Older Adults with the Space Fortress Game: The Role of Training Instructions and Basic Motor Ability," *Frontiers in Aging Neuroscience,* vol. 2, no. 145, 2010.

[22] Shebilske, W., Goettl, B., Regian, J.W, "Executive control of automatic processes as complex skills develop in laboratory and applied settings," in *Attention and performance XVII: Cognitive regulation of performance: Interaction of theory and application*, Cambridge, MA, The MIT Press, 1999, pp. 401-432.

[23] Nikolaidis, A., Goatz, D., Smaragdis, P., Kramer, A., "Predicting Skill-Based Task Performance and Learning with fMRI Motor and Subcortical Network Connectivity," in *International Workshop on Pattern Recognition in NeuroImaging (PRNI)*, Stanford, CA, USA, 2015.

[24] Goodfellow, I, Bengio, Y., and Courville, A, "Deep Learning," 2016. [Online]. Available: http://www.deeplearningbook.org. [Accessed 10 3 2017].

[25] Koch, G., Zemel, R., Salakhutdinov, R. , "Siamese Neural Networks for One-shot Image Recognition," in *International Conference on Machine*, Lille, France, 2015.

[26] Vezhnevets, A.S., Osindero, S., Schaul, T. Heess, N., Jaderberg, M., Silver, D., "FeUdal Networks for Hierarchical Reinforcement Learning," *ARXIV,* p. https://arxiv.org/pdf/1703.01161.pdf, 2017.

[27] Graves, A., Wayne, G, Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., "Hybrid computing using a neural network with dynamic external memory," *Nature,* vol. 538, p. 471–476, 2016.

[28] Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., van den Driessche, L., et al., "Mastering the game of Go with deep neural networks and tree search," *Nature,* vol. 529, pp. 484-489, 2016.

[29] Frederiksen, J. R., White, B. Y., "An approach to training based upon principled task decomposition," *Acta Psychologica,* vol. 71, pp. 89-146, 1989.

[30] Matiisen, T., "Simple deep Q-learning agent," github, 2016. [Online]. Available: https://github.com/tambetm/simple_dqn. [Accessed several times 2016].

[31] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., Zaremba, W., "OpenAI Gym," 2016. [Online]. Available: https://arxiv.org/pdf/1606.01540.pdf . [Accessed 11 3 2017].

[32] Google Deep Mind, "Code for Human-Level Control through Deep Reinforcement Learning," 2016. [Online]. Available: https://sites.google.com/a/deepmind.com/dqn/. [Accessed 15 12 2016].

[33] B. Schölkopf, "Learning to see and Act," *Nature,* vol. 518, p. 486–487, 2015.

[34] Miyoshi K., "Asynchronous Methods for Deep Reinforcement Learning," 2017. [Online]. Available: https://github.com/miyosuda/async_deep_reinforce. [Accessed 20 03 2017].

This page is intentionally left blank.

NLR

Anthony Fokkerweg 2
1059 CM Amsterdam, The Netherlands
p) +31 88 511 3113 f) +31 88 511 3210
e) info@nlr.nl i) www.nlr.nl