DRDC-RDDC-2015-P133

Professional Soldier Assessment of a Rifle-mounted Target Hand-off System

Jerome Levesque, Katherine Banko
Defence Research & Development Canada
Ottawa, Canada
Jerome.Levesque@drdc-rddc.gc.ca

Olaf Binsch
Netherlands Organization for Applied Scientific Research
Soesterberg, The Netherlands
Olaf.Binsch@tno.nl

ABSTRACT

The miniaturization of digital image acquisition and processing hardware, positional sensors, and batteries has enabled the creation of assisted targeting systems light enough to be integrated onto small firearms to increase the probability of soldiers detecting and hitting targets. As well, the technology allows soldiers to share target locations, thereby increasing tactical situational awareness and enabling target prioritization and target hand-off. We investigated how these new technologies might impact operational effectiveness by testing the concepts using human-in-the-loop simulation in a virtual environment. Two conditions examined the tool usage (no target hand-off vs. target hand-off). Within these conditions we added patrol and attack variants (no enemy, inaccurate enemy and accurate enemy). Each condition was repeated 8 times for a total of 64 randomized trials. Combat effectiveness measures quantifying blue casualties and the disruption of enemy activity were augmented with physiological indicants of stress and self-report measures of self-efficacy, performance and cognitive load. Null hypothesis significance testing applied to the combat effectiveness measures did not detect any statistically significant improvement in the combat effectiveness of the section as a result of using the target hand-off system. A Bayesian analysis was conducted to determine the probable size of an undetected effect. The human factor measures indicated differences between the simulated high and low threat conditions. Self-report measures combined with physiological measures did not reveal increases in stress when high and low levels of threat were compared. While participants evaluated the target hand-off system positively, the ability of the new technology to decrease cognitive load and therefore increase combat effectiveness measures remains unconfirmed. Simulations have limitations, particularly when exploring the benefits of target hand-off functionality (i.e. weapons effects and risks encountered in combat cannot be fully represented for safety and ethical reasons). And, combat stress is difficult to produce in an experimental setting. However, despite the small number of participants (n = 8), it was possible to estimate the probability distribution for the actual effect size.

ABOUT THE AUTHORS

Dr. Jerome Levesque is an operations research analyst with Defence Research & Development Canada. His work with the Canadian Army involves the development and use of combat models to provide advice for equipment acquisition and management programs. He obtained his Ph.D. in Atomic, Molecular and Optical Physics in 2006.

Dr. Katherine Banko is an operations research analyst with Defence Research & Development Canada. She has expertise in operations assessments, social science research methodologies, survey design, survey tools, seminar wargames, subject matter expert knowledge elicitation and quantitative tests and measures. She has a Ph.D. in Experimental Social Psychology and M.Ed. in Educational Psychology – Human Learning and Performance.

Dr. Olaf Binsch is a human factors scientist at TNO, Expertise Center of Human Factors in Soesterberg, The Netherlands. Before his engagement at TNO he was commanding officer of an infantry unit and military lecturer at the academy for junior leaders in the German Army. His research focuses on the interplay between social-psychology, psychophysiology, and physiology and he conducts experiments designed to monitor mental and physiological resiliency, and the physical, cognitive and social demands of military burden.

Professional Soldier Assessment of a Rifle-mounted Target Hand-off System

Jerome Levesque, Katherine Banko
Defence Research & Development Canada
Ottawa, Canada
Jerome.Levesque@drdc-rddc.gc.ca

Olaf Binsch
Netherlands Organization for Applied Scientific Research
Soesterberg, The Netherlands
Olaf.Binsch@tno.nl

INTRODUCTION

The miniaturization of digital image acquisition and processing hardware, positional sensors, and batteries makes it possible to create assisted targeting systems light enough to be integrated onto the sights of small firearms. Within the Canadian Future Small Arms Research (FSAR) project and the Dutch V1135 research programme "Next Generation Small Arms Systems", Defence Research and Development Canada (DRDC) and the Netherlands Organization for Applied Scientific Sciences (TNO) are investigating the potential to improve rifleman performance within the next 10 to 15 years using such technologies.

In the context of this experiment, target hand-off is a process used by dismounted infantry riflemen to coordinate the engagement of dismounted targets. The main treatment consisted of providing participants with a simulated system to assist in that process. The system allowed soldiers to designate the location of targets or objects of interest and make their location visible to other shooters through a digital display integrated onto each weapon sight.

The objective of this study was to estimate changes in combat effectiveness as a result of using a target hand-off capability integrated onto small arms sights and to assess the effects of cognitive load and stress on shooter performance. We hypothesised that using a rifle-mounted target hand-off system would result in increased effectiveness of the dismounted combat section, an increase in cognitive load but with a reduction in riflemen stress.

Because of time constraints under which the experiment was conducted and the small number of participants, the context was narrowed to specific environmental and engagement conditions. The scenarios would only be executed in daytime and in clear weather conditions. Two types of scenarios, urban patrol and attack, were chosen in order to provide different engagement intensities that cover a wide enough scope to test the technology, representing likely situations that dismounted soldiers might encounter in future operational contexts. While several useful conclusions are drawn from the results, the small number of participants necessarily increases the uncertainty of effect sizes. For combat effectiveness measures, we were able to estimate this uncertainty through Bayesian methods.

METHOD

Participants

Thirteen infantrymen volunteered and gave informed consent to participate in the experiment. The section consisted of one Lieutenant, one Sergeant, one Master Corporal, five Corporals and five Privates. They were divided into different roles and played the same position assigned throughout the experiment; Red Force (RF; n = 4) and Blue Force (BF; n = 8). The Lieutenant acted as both the platoon leader and oversaw RF activity (n = 1). Of the BF players, six played the roles of riflemen numbered as, RM1, RM2 and so on, one acted as the section commander (Comd), and one played the role of the second in command (2IC).

The participants served on average 64 months (SD = 50; range: 15 – 183 months) with the Canadian Armed Forces. During their military service, six participants (46%) were deployed at least once. Ten participants (77%) reported that they played video games in their spare time for on average 5.7 hrs/wk (SD = 4.), 3.0 hrs/wk (SD = 1.5); of those who played, 40% considered himself to be a 'serious' gamer (n = 4).

Setup

Each participant operated a desktop Virtual Battlespace 2 (VBS2) station. Those playing BF were paired in fire teams but separated by mobile walls; only BF used the target hand-off capability. Four others played either RF or, when scenarios included less than four enemy, civilian avatars. These four participants were directed by the Lieutenant, who acted as both the platoon leader and oversaw RF activity or supervised civilian roles as required by the scenarios. RF role players were situated away from BF players to avoid the chance of overhearing their communications. In addition to the civilians (n<5) controlled by RF players, 300 civilians were simulated in patrol scenarios and 25 in attack scenarios.

Participants wore a headset with a microphone for voice communications. BF used an intercom channel which was open throughout the missions and did not require manual operation in order to reproduce the functionality of Personal Role Radios (PRR). The section commander and the 2IC had access to another channel to communicate with the platoon commander. Depending on the parameters of each mission, BF had the ability to use a simulated target hand-off system, integrated in VBS2.

Figure 1 shows the VBS2-compatible target hand-off simulation created for the experiment. Each BF participant in the simulation had the ability to designate a point in space (be it on a human or an inanimate object such as a building) and broadcast the location to other team members. When a point location was broadcasted, information became visible to other teammates in their weapon sight. If a shooter was not facing the appropriate direction with respect to the target, a red arrow appeared suggesting a rotation direction towards the point being broadcasted. If the target was in sight, a flashing red diamond indicated the estimated location of the target. When sensor errors and network lag were set to zero, which was the case throughout this experiment, the red diamond mark indicated the exact location of the point that was broadcasted.

Figure 1. Target hand-off example with two players positioned on opposite sides of a bridge. Left panel: Player 1 designates ("lases") the target (small green square at the center). Right panel: Player 2, located out of sight of Player 1, receives information about the target location in his weapon sight (red arrow on the left).

Dependent Measures

Target Hand-off System Usage.
Each time a point in space was broadcasted using the target hand-off system an event was logged and tagged with mission type, user name, start time and duration. While this allowed us to capture every broadcast data, effective usage by receivers could not be monitored since reception happened in a passive manner (i.e. every broadcast is always displayed to all team members). This usage was reflected in the self-report questionnaires, however.

Team Combat Effectiveness Measures.
Two measures were used to quantify combat effectiveness: the number of casualties within the section (blue casualties) and the lifetime of enemy shooter, which was defined as the duration between each enemy's first shot and their incapacitation. By design, blue casualties could only occur in attack scenarios with accurate enemy fire

(25% of the missions). For realism it was necessary to limit the rate of blue casualties by appropriately adjusting the enemy's killing effectiveness. Enemy lifetimes were measured for all enemy entities (two to four per mission, for 75% of missions). This larger ensemble of values provided more statistical power than the number of blue casualties.

Individual Human Factors Measures.
Questionnaires were created to assess participants' attributions of performance, group and individual evaluations of performance, personal readiness, self-efficacy, feelings about the validity of the scenarios, and to provide information about the sight. All items were rated using 10-point Likert-type rating scales anchored with 1 = lowest rating and 10 = highest rating. Cognitive load was measured using the NASA TLX® workload index (Cao, Chintamani, Pandya, & Ellis, 2009). Physiological measures consisting of heart rate, respiratory rate, and galvanic skin response (GSR) were collected as indicators of stress and work load.

Experimental Design and Description of Scenarios
The experiment employed two main conditions to examine the tool usage (baseline: no target hand-off vs. experimental: target hand-off). Within these conditions, patrol and attack variants (without enemy, inaccurate enemy and accurate enemy) were used in order to examine the capability of the target hand-off system (see Figure 2). All conditions were randomized throughout the experiment.

The patrol and attack scenarios differed in duration which was dependent on how long it took participants to complete them; we recorded that the patrol scenario lasted on average 15 minutes and the attack scenarios lasted about 10 minutes. They also differed with respect to the number of civilians present. About 300 civilians were simulated in the patrol condition and between 15 and 25 civilians were simulated in the attack condition depending on the layout of the terrain (e.g. location of buildings). For the patrol scenarios, there were two conditions: no RF (enemy) present and RF present (n = 4) with low shooting accuracy. This latter condition was intended to create the perception to the BF players that the BF avatars could be killed by the enemy in order to keep the simulation as realistic as possible for the participants, while allowing us to tightly control the experimental manipulations. In actuality, there would be no BF casualties. As well, enemy fire would expose the RF locations allowing BF the opportunity to use the target hand-off. During the RF attack scenarios, RF (n = 4) fired accurately and inaccurately. We repeated each condition 8 times for all conditions for a total of 64 trials.

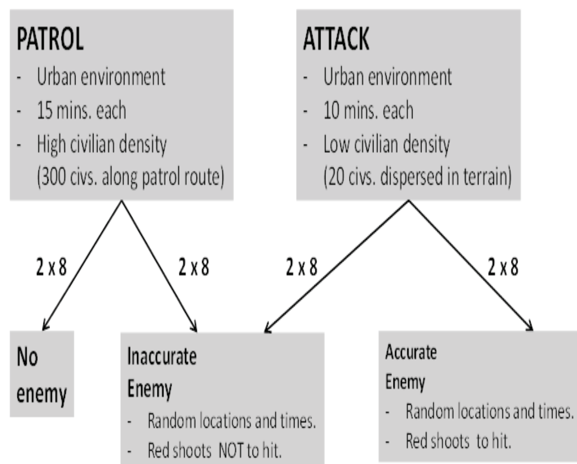

Figure 2. Experimental design depicting the patrol and attack conditions and their variations in non-enemy, accurate and inaccurate enemy forces used during the scenarios.

Because it was also interesting to compare the context (patrol vs. attack), in order to reveal in which context the target hand-off system would be more beneficial, the level of context was added into the study design. As the level of threat is normally different between patrol and attack context, three increasingly changing threat levels were introduced. That is, (1) none threatening situation (i.e. only civilians present in the scenario), (2) inaccurate enemy (i.e. armed RF shooting inaccurately in the scenario) and (3) accurate enemy (i.e. armed RF shooting accurately with the intention to kill BF). As the patrol context is more likely to deal with the first two threat levels and the attack context with the latter threat levels, those two threat levels were established for patrol and attack scenarios, respectively. In sum, there were 2 levels of condition (baseline vs. experimental) x 2 levels of context (patrol vs. attack) x 2 levels of threat (low vs. high) (see Figure 2).

Procedure

On the first day, participants were told the purpose of the study; informed consent was obtained. After role assignment, a training session was undertaken which included an explanation of the keyboard and the keystrokes needed to move the avatar and manage the target hand-off functionality. Prior to playing the different patrol and attack scenarios, they were briefed by the Lieutenant and the section commander about the content and procedures for each scenario. Following 'game' play, participants discussed the Techniques, Tactics and Procedures (TTPs) of each scenario. Participants assigned to the platoon section had two sensors taped to their non-dominant hand for measuring GSR. A chest belt system was strapped around their torso which housed a battery, data collector, GSR, respiratory and heart rate sensors.

Simulation Output Data

The data collected consisted of all combat events, time-stamped and included information about user id, point of origin and point of impact for each fire event and duration for each target location broadcast when target hand-off was used.

Combat Effectiveness

Two measures were used to quantify combat effectiveness: number of BF casualties and the lifetime of enemy shooters. The use of RF lifetimes as a combat effectiveness measure relies on the simple reasoning that as the section becomes more effective, enemy shooters are incapacitated in shorter times. Reduction of enemy lifetimes is typically correlated with reduction of engagement durations and BF casualties in probabilistic models of attrition in combat (Washburn, & Kress, 2009).

By design, blue casualties could only occur in attack scenarios with accurate enemy. Consequently, there were only 2 sets of 8 repetitions available for comparing the baseline condition with the use of target hand-off. Enemy lifetime on the other hand was defined as the time interval between an enemy shooter's first shot and the time of incapacitation. If enemy lifetime was reduced as a result of the BF using a new capability it would therefore be considered as having a beneficial consequence on combat effectiveness. Enemy lifetimes were observed for each enemy entity in the 3 out of 4 mission types where enemy were present, 2 sets of 80 points each, about 10 times more than the number of points available to compare blue casualties.

Human Factors Data

The self-report data was collected with an online survey tool; a physical monitor measured heart rate, galvanic skin response, respiration frequency and skin temperature. Attributions and evaluations of performance, perceived technology usage and cognitive load were collected following each scenario. Twice per day (before lunch break and at the end of the day), personal readiness and self-efficacy assessments were taken. At the end of each day, information about the validity of the tasks and the scenarios were provided. The data from the physical monitors, stored on chest belt systems, were prepared for the analysis using MATLAB. Finally, the participants assessed the new technology. Physiological measures were collected daily.

Average respiration cycles were measured in a defined time frame of 6 minutes for each participant as a function of condition, context and threat. This was selected by examining all scenarios to determine the shortest length of time to completion, which turned out to be 9 minutes. Generally, data from the first 2 minutes and the last minute were either missing or were incomplete in the recorded respiration data revealing a 6-minute period that was available for reliable analysis. Therefore, the data from the 'middle' was determined as the most reliable respiration cycle for all measured scenarios with the intent to compare all respiration data between the different conditions.

Periodically, group feedback was solicited to inform on-the-spot minor modifications to the scenarios (e.g. number of civilians present during a scenario). On the final day, general feedback about the activity and the logistics was collected in an open discussion format preceding the participant debriefing.

The human factors data were analysed with one-way analyses of variance (ANOVA) or repeated measures ANOVAs, depending on the situation. To analyse the different human factor variables and survey dataset, a 2 levels of condition (baseline vs. experimental) x 2 levels of context (patrol vs. attack) x 2 levels of threat (low vs. high) was used primarily. Participants' perceived usage of the technology was assessed with repeated measures ANOVA

(2 levels of condition x 4 levels of context (patrol, attack x [share targets, assign targets, designate reference points, designate civilians]). Finally, one-way ANOVAs (context x specific items/factors) were conducted to assess the evaluation of the technology and performance attributions; pair-wise comparisons (t-tests) were conducted to identify specific mean differences when applicable.

RESULTS

The usage of target hand-off for broadcasting point locations was cumulated for all users in each mission repetition, giving a single usage figure (in seconds) per repetition. A Kruskal-Wallis test provided a *p*-value of 0.13, meaning that within the null hypothesis, the probability of a discrepancy equal or larger occurring at random was 13%.

Table 1 shows the total usage for each role in the section, averaged over all missions. The usage numbers for C7 and C9 riflemen are given as averages per soldier. The results indicated that the section 2IC used the system the most; the commander used it the least. Note that these numbers only reveal how much each soldier used the system to designate targets. It was not possible to quantify how much the broadcasted locations were in turn used by section members, since this information was received passively.

Table 1. Average target hand-off usage (seconds per mission) for each engagement type and each section role.

Section role	Patrol	Attack	All
Codr	0.0	0.6	0.3
2IC	18.1	29.3	23.7
C7	4.4	5.4	4.9
C9	3.8	3.2	3.5

Perceived Technology Usage

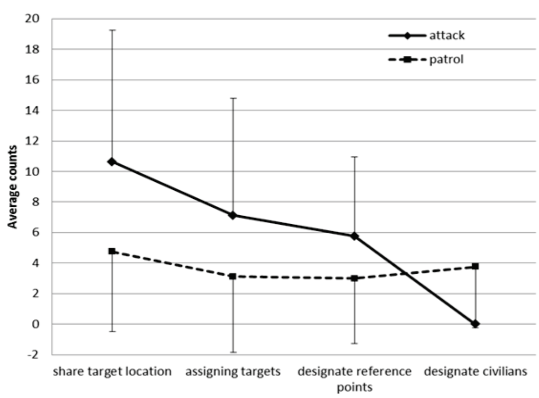

Figure 3. The interaction effect of condition and context.

A 2 (condition: baseline, experimental) x 4 (context: patrol, attack x [share targets, assign targets, designate reference points, designate civilians) repeated measures ANOVA was conducted using the technology usage dataset revealed a significant main effect of condition, $F(1, 7) = 6.9$, $p < .05$, a main effect of context $F(3, 21) = 5.7$, $p < .01$, and a significant interaction effect between context and condition ($F(3, 21) = 7.1$, $p < .01$). Post hoc pair-wise comparisons revealed that the participants who used the new technology more often for sharing targets, assigning targets to someone else and designating reference points during the attack context (on average 3 more times; compared to the patrol context), but did not use the new technology to designate civilians.

Combat Effectiveness Measures

Blue Casualties.
By design, casualties inflicted on the BF section could only occur in attack scenarios with accurate enemy. Consequently there were only 2 sets of 8 repetitions available for comparing the baseline condition with the use of target hand-off. The numbers of casualties were respectively (2, 2, 2, 2, 2, 2, 1, 2) for the baseline cases and (2, 2, 1, 3, 2, 2, 2, 3) for the target hand-off cases. A Kolmogorov-Smirnov test on these two sets gave a p-value of 0.71. A p-value this high indicated no detectable effect of using target hand-off in reducing the probability of blue casualties. The BF casualty measure had relatively low statistical power because of the small sample available and also because the values were small integers, which resulted in low resolution.

Enemy Lifetimes.
Our first approach was to analyze RF lifetime data using Null-Hypothesis Significance Testing (NHST). A series of Kolmogorov-Smirnov (KS) tests comparing RF lifetime distributions in the baseline and target hand-off conditions gave p-values of 0.34, 0.33 and 0.75 respectively for patrol with inaccurate enemy, attack with inaccurate enemy

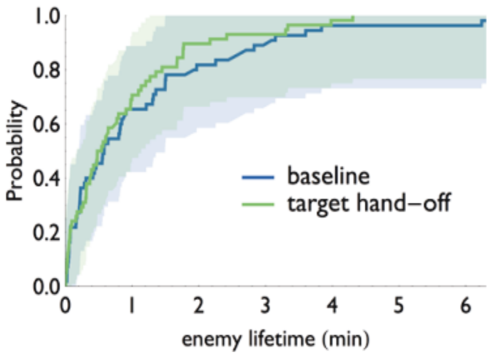

Figure 4. Cumulative distribution functions for enemy lifetimes (showing 95% error bands).

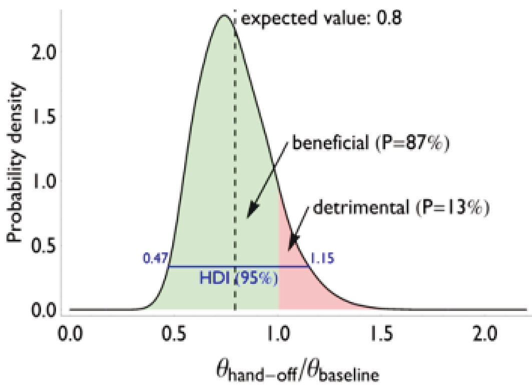

Figure 5. Probability distribution for the ratio of expected RF lifetimes in hand-off vs. baseline conditons.

and attack with accurate enemy. Therefore, on a per-scenario basis, there were no significant changes in the RF lifetime distributions as a consequence of using target hand-off. In order to increase statistical power this comparison was repeated without discriminating over mission type, resulting again in failing to reject the null hypothesis (p-value=0.68). Figure 4 shows the cumulative probability distributions for RF lifetimes, all mission types combined, along with the 95% confidence interval bands as determined using the KS statistic. The two cumulative distributions are within each other's error bands.

The NHST approach addressed the following question: assuming that an effect is *absent*, how likely is the difference between the samples? A significant effect is then postulated only if the p-value was sufficiently small. Failure to reject the null hypothesis might have two roots: the absence of an effect, or an effect small enough that it cannot be resolved within the statistical power of the experiment. The latter is a concern for small-scale experiments such as the one performed here. The test does not distinguish between these two causes. The following, different question remains unresolved: assuming that an effect is *present*, what do the experimental results say about its probable magnitude? This question can be resolved by Bayesian analysis methods. The analysis presented here allowed us to leverage the results of this small scale experiment and obtain a probability distribution for how enemy lifetimes may vary as a consequence of using target hand-off.

The first step in the Bayesian analysis is to find a credible model for the RF lifetime data shown in Figure 4. A decreasing exponential model was rejected by a KS test but a gamma distribution model was accepted (p = 0.84 and 0.93 respectively for baseline and hand-off). Gamma distributions are specified by two parameters: k (shape) and θ (scale). The best fits for baseline and hand-off data (Figure 4) have similar values for the shape parameter k (0.65 for baseline vs 0.68 for hand-off) but values further apart for the scale parameter θ: 1.7 for θ_{baseline} vs 1.3 for $\theta_{\text{hand-off}}$. Consequently we chose as the likelihood function a gamma distribution with k fixed at 0.66, and a variable θ. This likelihood function was used in conjunction with a uniform prior to calculate the joint posterior distribution for θ_{baseline}, and $\theta_{\text{hand-off}}$. This result was then used to calculate the probability distribution for the ratio $\theta_{\text{hand-off}}/\theta_{\text{baseline}}$, which indicated how much shorter (if smaller than 1) or longer (if greater than 1) the expected enemy lifetime would be when using target hand-off. This probability distribution is shown in Figure 5. The experiment tell us that assuming an effect is present, there is an 87% chance that it is beneficial. The most probable effect of using target hand-off at the section level would be a reduction of about 20% in enemy lifetimes. There is uncertainty in the estimate, as reflected by the width of the curve. The 95% Highest Density Interval (HDI) of the distribution shown in Figure 5 stands between 0.47 and 1.15, a width of 0.68.

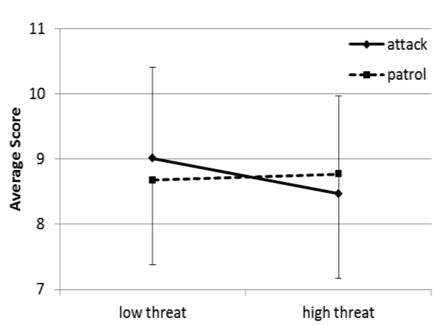

Figure 6. Interaction of scenario and threat.

Perceived Group Performance

The 2 (condition: baseline, experimental) x 2 (context: patrol, attack) x 2 (threat: low – civilians or inaccurate enemy fire, high – inaccurate or accurate enemy fire) repeated measures ANOVA on the perceived group performance showed a significant interaction effect between context and threat $F(3, 21) = 6.6$, $p < .05$ shown in Figure 6. Post-hoc pair-wise comparisons on this interaction revealed that participants rated the group performance on average higher during the attack context when the threat was low (i.e. civilians present or inaccurate fire from enemy during the patrol and attack contexts respectively) compared to the high threat condition (i.e. inaccurate or accurate fire from the enemy during the patrol and attack contexts, respectively); whereas the estimation of the performance during the patrol session remained the same for both high (i.e. inaccurate fire from enemy) and low (i.e. civilians) threats.

Attributions of Performance

The one-way ANOVA on the 9 (attribution) performance factors revealed a significant main effect of attribution, F (8, 56) = 8.1, $p < .01$. Post hoc pair-wise comparisons showed that the participants attributed their performance to their operational skill level, training experience, confidence, team members, and the weapon technology, but not to external pressures (ps < .0).

NASA TLX Work Load Index

The 2 (condition: baseline, experimental) x 2 (context: patrol, attack) x 2 (threat: low, high) repeated measures ANOVA on the weighted workload index revealed no main or interaction effects, all F s < 2.0, $p > .1$.

NASA TLX Mental Workload

The 2 (condition: baseline, experimental) x 2 (context: patrol, attack) x 2 (threat: low, high) repeated measures ANOVA on the percentages of the cognitive workload (as one out of six TLX workload components) revealed a significant main effect for context, F (1, 7) = 7.3, $p < .05$ and a significant interaction between condition and threat F (1, 7) = 19.9, $p < .01$) shown in Figure 7.

Post-hoc pair-wise comparisons on the main effect revealed that the participants rated their mental demands during the attack context 5.7% higher compared to the patrol context. The pair-wise comparisons of the interaction showed that perceived mental workload during the high-threat baseline condition was low. In contrast, perceived mental workload was higher during the experimental conditions (i.e., when they are using the new technology) and when the threat was high (3% difference between these end points, $p < .05$).

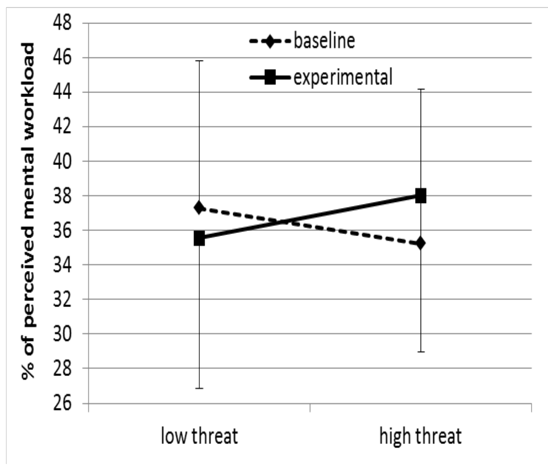

Figure7.The interaction of condition under high and low levels of threat.

TLX Temporal Demand

The 2 (condition: baseline, experimental) x 2 (context: patrol, attack) x 2 (threat: low, high) repeated measures ANOVA on the percentages of the temporal workload revealed no main or interaction effects, all F s < 3.1, $p > .1$.

Physiological Measures

The 2 (condition: baseline, experimental) x 2 (context: patrol, attack) x 2 (threat: low, high) repeated measures ANOVA on the heart rate (beats per minute) revealed no main or interaction effects, F s < 3.1, $p > .1$.

Perceptions of the Technology

The one-way ANOVA conducted on the 11 evaluation factors revealed a significant main effect of evaluation, F (10, 70) = 3.4, $p < .01$. Post-hoc pair-wise comparisons showed that the participants found the (simulated) target hand-off technology a) useful for locating targets, b) beneficial for the section to locate targets and c) enhanced their combat effectiveness ($p < .01$). Obviously, some participants also used the information provided by the sight for navigation and found this sight capability of assistance ($p < .05$).

Self-efficacy and Personal Readiness

The one-way ANOVA conducted on the 11 factors revealed a significant main effect of self-evaluation, F (9, 63) = 18.3, $p < .001$. Post-hoc pair-wise comparisons on this effect showed that the participants felt energetic, capable, confident, calm and relaxed, but also tired ($p < .01$).

Scenario Ratings

The one-way ANOVA conducted on the 11 scenario factors revealed a significant main effect, $F(10, 70) = 15.5$, $p < .001$. Post-hoc pair-wise comparisons on this effect showed that the participants found the scenarios to be realistic, interesting, enjoyable, albeit somewhat easy, boring and tedious ($p < .0$).

DISCUSSION

The use of simulation to understand the behavior of something without testing it in real life is a well-accepted practice (NSF 2006). However, simulations have limitations, particularly when exploring the benefits of target hand-off functionality. The inability to create realistic battlefield conditions and measure the associated stress with knowing one's life is at risk is a challenging endeavor in simulation, implying that rather low effect sizes on stress measures are expectable. First-person 'shooter games' are also known to provide reduced tactical awareness in comparison to reality (Whitney, Temby, & Stephens, 2013), due to diminished visual and auditory cues. As a result, any tool enhancing tactical and situational awareness in VBS2 has the potential of having a greater effect than it would have in the field. That is, the simulation-based trial would likely give an upper bound to the system's contribution to increasing tactical awareness, which in turn positively benefits overall combat effectiveness.

Scenario Validity

As a check of the validity of scenarios, participants responded to 11 items to tap into the realism of the patrol versus attach conditions. The participants reported that the scenarios were realistic, interesting, enjoyable, yet easy, boring and even tedious. The latter findings accurately reflect the overall work of a soldier at times when not in battle (suggested by participants during the debriefing) whereas the former adjectives suggest that the scenarios were realistic enough to test the technology and that the participants were sufficiently motived throughout the experiment.

Target Hand-off Usage

Participants suggested that the target hand-off system should likely be implemented in a head-mounted device separate from the weapon sight, to avoid pointing weapons at people in populated areas or revealing the section's intentions ahead of time in high threat environments. Note that while attack scenarios in the experiment included four enemies and patrols two enemies, target hand-off usage did not change significantly between mission types.

Table 1 provides information on target hand-off broadcasts as a function of section roles. Broadcasts were unequal among section members, with the section 2IC accounting for over 70% of the total usage and the section Cmdr. less than 1%. This reflects how the section used the system in a tactical situation. In combat situations, a detachment of four soldiers including the 2IC would form a fire base to support the other four soldiers forming the assault team. Typically the 2IC would designate targets while the assault team engaged them with support from the fire base. In that sense, the 2IC was mostly providing target locations to the rest of the section.

Participants perceived they used the tool more during the attack scenarios to share target locations, assign targets to others, and to designate civilians compared to patrol scenarios with the exception of not using it to designate reference points. Within subject comparisons revealed that the Cmdr. and his 2IC used it the most regardless of condition or context. Notably, with the exception of the Cmdr., it was infrequently used to designate civilians in either the patrol or the attack context. Interestingly, while the Cmdr. accounted for only 1% of target broadcasts he nevertheless perceived that he was using the system at a high rate. This might be because he was mostly using target broadcasts made by his men to follow the tactical situation. Because the reception of target broadcasts was done passively, it could not be tracked in the database but it would have been reflected in the survey, which corresponds to what is observed.

Combat Effectiveness

Taking an approach based on Null-Hypothesis Significance Testing (NHST), Kolmogorov-Smirnoff tests did not confirm any effect of target hand-off on BF casualties or the distribution of enemy lifetimes. While there might be an absence of effect, it is also possible that an existing effect was too small to be detected with the sample sizes available. There might also have been a reduction in contrast due to the counter-insurgency context of the scenarios. RF shooters were sometimes difficult to locate immediately in densely populated environments. Such concealment gives a first shot advantage to the enemy that cannot be mitigated by a target hand-off capability (i.e. a target unknown to all shooters cannot be shared).

To answer how large the effect could be *if* there were any effect, we conducted a Bayesian analysis of the data. It was possible to derive a probability distribution for the reduction of enemy lifetimes due to target hand-off usage. Overall, our experiment indicates an 87% chance that target hand-off reduces enemy lifetimes by some amount. The most likely effect is a 20% decrease of enemy lifetimes as a result of using target hand-off. The uncertainty of the estimate is relatively high however, with the 95% HDI running from a 53% decrease to a 12% increase. It is estimated that a dataset four times the size would narrow the 95% HDI to a width of about 0.30. A dataset 16 times larger would reduce the 95% HDI to an estimated width of 0.20.

Using Bayesian data analysis provides a complete probability distribution for the effect under consideration, and is not bound by yes/no outcome of null hypothesis statistical tests. As an example, consider an experiment conducted with small samples. If the effect size is too small, it is likely that significance tests will turn out negative, even if an effect is actually present, simply as a result of insufficient statistical power. In other words, the frequentist approach is unlikely to provide any information in situations where the signal to noise ratio is low. However, Bayesian inference will still provide a probability distribution for the effect, provided there is a mathematical model for the phenomenon under observation. Notably, the lower statistical power of the experiment results in a relatively broad (i.e. uncertain) probability distribution. Nevertheless this analysis allows us to obtain additional information that is not provided by NHST.

Any benefits estimated here are most likely overestimated in comparison to what they would be in reality, for at least two reasons. It should be expected that sensor noise and network, which were not represented in this experiment, would reduce the effectiveness of target hand-off. Secondly, first-person 'games' are known to provide reduced tactical awareness in comparison to reality (Whitney, et al., 2013). As a result, any tool enhancing situation awareness in VBS2 has the potential of having a greater effect than it would have in the field.

Performance ratings

There were no differences found with respect to individual performance. Regardless of condition (baseline vs. experimental), participants reported positively on their performance independent of the new technology. For perceived group performance, there was a significant interaction effect; during the attack scenarios, perceived performance of the group decreased as the level of threat increased (Jones & Harris, 1967) providing some evidence that the threat manipulation which was intended to induce higher levels of stress resulted in perceptions of poorer performance, a finding in line with the U.S. Army's Advanced Combat Rifle findings (Radcliffe, 2008).

Attributions of Individual Performance

When asked to report on the source of their performance, participants positively attributed their individual performance to their operational skill level, training experience, confidence, other team members, and the weapon technology. However, the weapon technology was rated lower, on average, relative to attributions of their soldiering skills. They did not attribute their performance to external pressures; that is, they did not attribute their individual performance to time pressure, situational pressure, or pressure due to being evaluated. This finding provides further evidence that the threat manipulation within the scenario was successful at inducing a stressful situation; that is, it was not due to demand characteristics from the experimental setting (Orne, 1962).

Mental Workload

Overall, there was evidence that cognitive burden increased when participants used the technology. The TLX scale was a composite that examined different workload domains. Because we hypothesized that the usage of the new technology would enhance combat effectiveness, it allowed for additional analyses for the human factor domains of mental workload, temporal workload and cognitive/sensory effort. Drilling down into these specific workload items, we found a mix of results. Although these results did not directly reveal the expected benefits, the results underscore the validity of the experimental set-up, as the participants perceived less mental workload during the patrol context compared to the attack context.

Furthermore, when the threat was high the perceived mental demand was also high and the usage of the new technology was perceived as extra mentally demanding in the attack context compared to the patrol context. It is unclear yet why the between-subject analyses of the mental, temporal and cognitive workload items revealed differences between different participants. It is likely that the participants perceived the conditions, context and/or threats differently, possibly because they had different roles, experiences, combat readiness or some combination of these.

Physiological Measures

There were no overall differences with respect to the skin conductance or skin temperature measures. There was a significant difference between-subject effect in heart rate and heart rate variability (bpm), and in rates of respiration. Specifically, the Cmdr. had significantly higher heart rate and respiration levels relative to the rest of his section. In addition, during the experimental conditions with tool usage, the section had higher respirations levels. Respiration increased significantly in the patrol condition under high threat. In the attack condition, the respiration rate remained high during the attack scenario under both high and low threat conditions. These finding compliments the perceived workload results. That is, the Cmdr. evidenced a physiological indication in keeping with his perception of experiencing a higher workload relative to all others in the section.

CONCLUSION

At the onset of this experiment, questions related to the benefits of a dismounted target-sharing capability on combat effectiveness had not yet been explored. Moreover the new simulation software created for modelling next-generation small arms in VBS2 had not been used in any other experiment as yet. In face of this relative uncertainty an attempt was made to minimize risk by keeping participation at the scale of a Canadian infantry section (8 soldiers). Several useful conclusions can be drawn (despite a reduced statistical power): Target hand-off is most likely beneficial for increasing group combat effectiveness (~87% likely), with an expected reduction of enemy lifetimes by 20%. These benefits might be upper bounds however: the addition of sensor noise and network lag might reduce the system's effectiveness. Also, because navigation in VBS2 is more difficult than in the field, aids to situational awareness might be over-effective in a gaming environment. Usage of target hand-off is role-dependent. The section 2IC performed most target designations, providing information to assaulters and the section Cmdr..

While VBS2 does not induce combat-like stress levels, a triangulation of measures was observed between physiological measures of stress and the intensity of simulated combat. This result indicates that game-based experimentation should not be ruled out as a tool for obtaining insight on how information technologies impact the soldier's cognitive load.

This study focused solely on dismounted patrol and attack missions. In future work it will be necessary to assess the impact of dismounted target hand-off in defensive scenarios. Practical limits on sensor noise and network lag will also need to be established.

Current studies of next-generation small arms concentrate on the benefits, limitations and costs of technological solutions. In the future it might be fruitful to conduct a corresponding analysis with respect to investments in training, with the perspective of optimizing the combination of technology with training.

REFERENCES

Cao, A., Chintamani, K., Pandya, A., & Ellis, R. D. (2009). NASA TLX: Software for assessing subjective mental workload. *Behavior Research Methods, 41*, 113-117.

Jones, E. E., & Harris, V. A. (1967). The attribution of attitudes. *Journal of Experimental Social Psychology, 3*, 1–24.

National Science Foundation (NSF) Blue Ribbon Panel (2006). Report on Simulation-based Engineering Science: Revolutionizing Engineering Science through Simulation. NSF Press

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776-783.

Radcliffe, Col. R. (2008). NDIA Small Arms Symposium, *presentation at the International Infantry & Joint Services Small Arms Systems Symposium, Exhibition & Firing Demonstration, Dallas (TX), 20 May 2008*. 11 slides. Recovered on 19 June 2014 from http://www.dtic.mil/ndia/2008Intl/Radcliffe.pdf.

Washburn, A. R. & Kress, M. (2009). Combat Modelling. Heidelberg: Springer.

Whitney, S. J., Temby, P. & Stephens, A. (2014). A review of the effectiveness of game-based training for dismounted soldiers. *Journal of Defence Modeling and Simulation, 11, 319-328.*