Towards a Translational Method for Studying the Influence of Motivational and Affective Variables on Performance During Human-Computer Interactions

Jason S. Metcalfe[1,2(✉)], Stephen M. Gordon[1], Antony D. Passaro[1,2], Bret Kellihan[1], and Kelvin S. Oie[2]

[1] Scientific Research Department, DCS Corporation, Alexandria, VA, USA
{jmetcalfe, sgordon, apassaro, bkellihan}@dcscorp.com
[2] Human Research and Engineering Directorate, US Army Research Laboratory, Aberdeen Proving Ground, Aberdeen, MD, USA
kelvin.s.oie.civ@mail.mil

Abstract. A primary goal in operational neuroscience is to create translational pathways linking laboratory observations with real-world applications. Achieving this requires a method that enables study of variability in operator performance that does not typically emerge under controlled laboratory circumstances; the present paper describes the development of such a paradigm. An essential aspect of the design process involved eliciting subject engagement without using extrinsic incentive (e.g. money) as a motivating stressor and, instead, tapping an appropriate intrinsic incentive (i.e. competitive stress). Two sources of competition were initially considered including one based on self-competition and another based on competition with another individual; ultimately, the latter approach was selected. A virtual competitor was designed to affect individual valuation of momentary successes and failures in specific ways and preliminary results revealed early indicators of success in meeting this goal. Discussion focuses on implications and challenges for future research using similar translational paradigms.

Keywords: Competitive stress · Affect · Motivation · Translational science

1 Introduction

Research on operational performance in defense contexts increasingly leverages the knowledge, methods, and technologies of modern cognitive neuroscience to improve decision making within complex, high-risk, time-limited circumstances [1, 2]. The future success of such efforts relies partly on the ability to outfit warfighters with lightweight, wearable sensor suites that, in concert with the application of advanced computational techniques, facilitate the inference of a variety of cognitive and/or emotional processes (i.e. attention control, decision-making, event appraisal) and states (i.e. stressed, frustrated, fatigued; [3]). Ultimately, to make full use of such technologies, these types of efforts must establish clear translational pathways linking laboratory

D.D. Schmorrow and C.M. Fidopiastis (Eds.): AC 2015, LNAI 9183, pp. 63–72, 2015.
DOI: 10.1007/978-3-319-20816-9_7

observations of operator state changes with real-world implications for the development of systems and methods that will improve operational effectiveness.

Traditional research in cognitive neuroscience has tended to focus on delineating changes in psychological constructs, such as mental fatigue or attention, using narrow sets of variables derived from physiological and behavioral data recorded during simplified tasks that were performed in tightly controlled laboratory environments. For instance, some studies assess brain activity that has been averaged across hundreds of isolated stimulus presentations using neuroimaging techniques that are limited in spatial and/or temporal resolution (e.g. functional near-infrared spectroscopy, fNIRS, or electroencephalography, EEG). While these paradigms have provided important understandings of average brain dynamics and behavioral response patterns across populations of individuals, they do not provide much power for understanding precisely how these patterns may vary within any given operator or from moment to moment during individual performance instances within more complex task and stimulus environments. A precise and individualized understanding, however, is necessary to enhance real-time, real-world operational effectiveness of systems and technologies intended to work for individuals in varied operational circumstances. New paradigms are thus needed to enable detailed understandings of brain-behavior relationships in individuals as they perform real-world operational tasks and make real-time appraisals thereof [4].

The overarching goal of the project described in this paper was to develop and validate methods for enabling and drawing inferences about event appraisal processes (i.e. determining the value of actions and outcomes) as reflected in motivational and emotional state variables. Specifically, we aimed to develop an understanding of factors that influence the event appraisal processes expected to be revealed by subjective changes in affect, task-specific motivation, and their associations with performance.

2 Method

2.1 Assumptions

The research described herein was designed based on certain assumptions about the information available for inference and classification of affective and motivational states within realistic operational environments. Even in highly constrained operational settings, such as tactical vehicle environments, there are sources of information that have considerable potential to improve the estimation of operator states, particularly when applied in combination with those already commonly used in neuroscience laboratories (e.g. [5]). Further, we believe that such information may be mined to help identify states with unique, yet consistent, neural, cognitive, and behavioral dynamics. For example, the use of computer-vision technologies that register graph topologies with individual facial features may be applied simultaneous with facial electromyography (EMG) and electrodermal activity (EDA) recordings to assess and reliably characterize operator affective variables during a variety of task conditions. The present research was based on a method that we developed to leverage these approaches through the capture of continuous streams of data related to individual facial expressions, head orientation, and manual behavior, EEG, EDA, ECG (electrocardiography) and eye motion tracking, while subjects performed operationally relevant tasks. In addition to the behavioral and physiological modalities, state data from common task interface devices (keyboards, response pads, vocal inputs, etc.) as well as environmental event data were also expected to enhance the performance of traditional machine learning-based state prediction and classification systems.

2.2 Primary Task

Subjects were asked to perform an object detection and classification task in a semi-realistic virtual environment that emulated a modern military-relevant urban setting. There were four object types and each required a button-press response. The objects appeared within moving imagery that represented the out-the-window view of a tactical vehicle driving through an urban environment. As the vehicle proceeded, the objects appeared in a randomized order at a variety of radial distances and eccentricities. The subjects did not have any control over the simulated vehicle's path.

For each 15-min condition, there were 300 object presentations divided among four types. The object types included a male figure holding a large gun (human threat), a male figure with no gun (human non-threat), a table that the subject could not see under (table threat), and a table that the subject could see was clear underneath (table non-threat). Objects each appeared for 1 s and the inter-stimulus interval was chosen from a uniform random distribution between 1 and 3 s. Subjects were responsible for reporting the threat status (threat/non-threat) of each object by pressing one of two buttons on a response pad. For each subject, threats (regardless of human or table) were to be indicated using one hand and non-threats with the other; the assignment of hand to stimulus category was balanced across the subjects.

3 Design Considerations

The main goal of this research was to examine motivational and affective influences over variability in operational task performance. Among the avenues towards achieving this goal in a way that provided translational value (i.e., having a clear pathway through which it may be applied in operational settings) it was essential to have a means of facilitating a sense of investment in the task outcome that was analogous to motivational sources in the operational environment. Typically, research aiming to provoke a sense of personal investment in the task has accomplished this by linking performance outcomes to extrinsic rewards; most commonly these rewards have been in the form of financial incentives. Here, our concern was within the domain of defense application. Though financial incentives are indeed "real-world" (i.e. most adults are familiar and have daily experience with valuing things in financial terms and, further, have some motivational orientation towards financial gain), their function as a motivator on the battlefield, where most decisions have a potential cost of human life or overall mission failure, is not as clear. Indeed, it seems clear that not all types of incentives are equally rewarding; the behavioral context is an important indicator of how a given decision or behavior is best valuated [6]. For the current research, it thus was considered appropriate to implement a different reward mechanism. In particular, we were interested in a reward mechanism that was more internally driven and, at least in terms of face validity, was more relatable to a defense context than would be financial gain. As such, we developed the translational paradigm around competitive challenge. There were a few crucial design considerations driven by the goals just described. These considerations included the definition of the basis for competitive reward, the design of a scoring function for the reward feedback, and the appropriate selection of the subject population given the reward paradigm used.

3.1 Competitive Reward Basis

In considering what it might mean to motivate participants with a so-called competitive stressor, the first question encountered was who, or what, should serve as the basis for competitive evaluation. That is, against whose performance would the subjects evaluate their own performance? Immediately, it was decided that actual human confederates, though most preferred [7], were not feasible for several reasons including cost, time, and experimental control. Thus, the remaining options necessarily involved the use of a computerized competitor. Nevertheless, even within the domain of computerized competition, the reward structure has several options. Most generally, rewards can essentially be with the self or with another. That is, the rewards can be chosen to focus the individual on evaluations of their current performance with respect to their own recent performance history (self-competition). Alternatively, the reward structure can be chosen to focus the individual on their current performance in comparison with that of another. Therefore, during initial development of the software that served as the basis for this study, two variants of the competitive reward were developed and overlaid on the exact same game.

One of the two game variants that was considered was based on the concept of an "Infinite Runner". In this competitive format, game evaluators played to achieve the highest score possible. There was no direct competitor. Instead, the competitive feedback was structured such that the cost of an incorrect response increased with each correct response. That is, each correct response would earn the player a flag and an uninterrupted sequence of flags was required to score points, which were the main objective. Importantly, an incorrect response would cost the loss of all current flags, thus requiring the subject to start over in accumulating to the next point value. Further rewards and protections were possible; for example, enough points could earn "shields" that would offer protection against loss of flags. Thus, there were different levels of achievement combined with different degrees of consequence for performance errors. Ultimately, game evaluators indicated that the complex scoring structure of this game, while more interesting, was less intuitive and thus less preferred for a single session of game play (i.e. it would take longer to learn).

The second game variant, which was ultimately selected for the formal experiment, was based on a model of more direct competition. That is, players were told that they were in direct competition with a virtual competitor (VC) that they could out-perform if they played well enough. Their score was displayed in direct opposition to that of the

VC. In reality, the score of this VC was driven by a scoring function that accumulated or lost points as a transformation on the recent performance history of the human player (algorithm described in Sect. 3.2 below). Research advantages of this paradigm became immediately apparent. Though the game evaluators reported that this variant was less interesting, it was more intuitive and therefore, they had a more immediate understanding of what was needed to "win" the game. Moreover, a degree of simplicity in experimental control was afforded by the use of the VC scoring function. That is, the function was tunable to achieve a specific level of competition in terms of the percentage of time that the VC was ahead of or behind the performance of the player.

3.2 Structuring the Scoring Function

In order for the competitive stimulus to function properly (elicit subject investment and engagement to the point of provoking affective state changes), it was not only important that the subjects perceived the presence of the competition but also that the competition was of a nature that it made sense to fully engage with the game. That is, the players of this game needed to judge that there was a legitimate chance they could win even if the apparent difficulty was high. Indeed, if and when subjects perceived that the game was not "winnable" or was otherwise "rigged", their personal sense of agency and investment in the task would diminish. This risk was apparent in the formal data sample when one subject noted "I lost competitive interest when I figured out it was a slingshot AI" (implying fixed outcome). Now, despite the fact that the subject was incorrect about the precise nature of the VC (it was not, in fact, a "slingshot AI"), the important thing was that this resulted in complete disengagement in the task in terms of personal investment in the outcome. Further, without a sense of personal investment, the competitive reward would fail in provoking the kinds of motivational and affective state changes that were actually of interest in this study.

Therefore, a primary task in the development of this translational paradigm was to define the VC such that it tracked with the performance of the subjects in particular ways, was tunable to particular levels of competitive challenge (see below), and that its functional connection to the subject's own performance was not immediately apparent. The subjects must also have had a fair chance to win, even when winning was more or less difficult. Several mechanisms were attempted through early development and evaluation phases, including tracking of a weighted running average of the subject's performance as well as a number of different variations on proportional-integral-derivative (PID) based feedback control schemes. Ultimately, a manually-tuned proportional-gain feedback control model was the basis of the scoring function, with specific VC scores determined by the control system plus a stochastic value added as a mask against detection by the subjects.

3.3 Subject Selection

Logically, when choosing any particular reward mechanism to affect motivational and emotional states, it is essential to ensure that the subjects actually value the rewards in the manner expected. For instance, during the initial design phase for the stimuli to be used in the current paradigm, subjects who were self-identified “gamers” were thought to be important to include as the game and competitive reward were evaluated. The rationale for focusing on gamers as system evaluators was that they would already maintain an intrinsic reward system based on the gain and loss of points and, therefore, they would be motivated by this assumed value system. In the initial informal system tests, a number of the evaluators who self-identified as both gamers and non-gamers played the game under the two variants of the competitive reward system. Ultimately, the responses to the game versions appeared independent of whether an individual was a self-identified gamer. However, all evaluators were also asked additional questions about their game-playing experiences and, as it turned out, the key variable that appeared to elicit differing perceptions was whether the evaluators self-identified as feeling competitive when playing the games. Moreover, not all evaluators who self-identified as competitive were also gamers. While no formal analyses were conducted on these early evaluations, the lesson was that the match of reward to personality traits was likely to be critical as expected. Moreover, because the manipulation targeted competitive challenge, it was clear that self-identified competitiveness was more important than gaming experience.

4 Translational Paradigm

Sixteen adult male participants[1] who self-identified as being competitive were recruited to participate in this experiment. Each subject played the game in two conditions, with each condition representing a different level of competitive challenge. Subjects were instructed to play the object detection and classification game with the primary goal of outscoring the VC. The subject’s score was clearly displayed along with the score of the VC and both were updated in real time. Immediate score feedback was shown per target, with the numeric value for each response displayed as a brief pop-up overlaid on the screen. Correct and incorrect responses were associated with specific sounds connoting success or failure. The sound cues were intended to help the subjects more readily evaluate their responses. To provide a persistent reminder of the competitive aspect of the game, the difference between the subject’s score and that of the VC was graphically represented with a color-coded bar at the bottom of the screen. The size, direction, and color of the bar changed as a function of the score difference to indicate whether the subject was winning (blue bar growing to the right of screen center) or losing (red bar growing to the left of screen center).

Although subjects were told that the other score that appeared on the screen was based on a computerized competitor (the VC), they were also told that the outcome of the game was not fixed. Though the outcome was not fixed, the score of the VC was controlled in a systematic manner relative to the subject’s real-time performance. To give the impression of two levels of competition, the VC either closely matched the subject (high challenge) or was clearly losing a majority of the time (low challenge). The actual

[1] All participants in this research provided their voluntary, fully informed consent [8, 9], and investigators adhered to U.S. Army policies for protecting human subjects [9].

difficulty of the task was comparable across conditions. Figure 1 shows distributions of the score differences between the subjects and the VC for each condition.

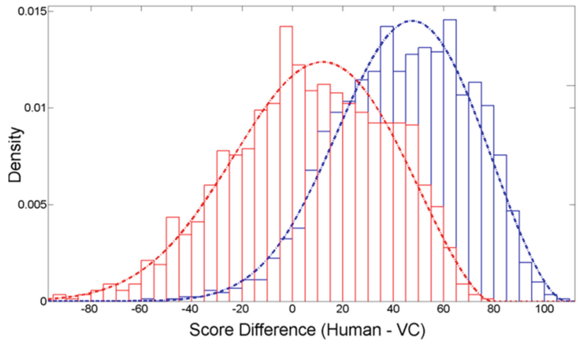

Fig. 1. Empirical distributions of the score differences between the subjects and the VC. The distribution centered on 0 points (red) was from the high challenge condition and the one centered on 50 points (blue) was from the low challenge condition. (Color figure online)

In addition to the level of persistent competitive challenge, the stimulus environment was varied to change the task difficulty within each condition. The intention of this manipulation was to shift the task context, within each competitive circumstance, in order to enable assessment of corresponding changes in affective state variables as a function of the ambient context. Here, the task difficulty was periodically increased with a fog that was overlaid onto the video stream. Shown in Fig. 2, these periods of fog (shaded regions) were nearly identical across the two challenge conditions.

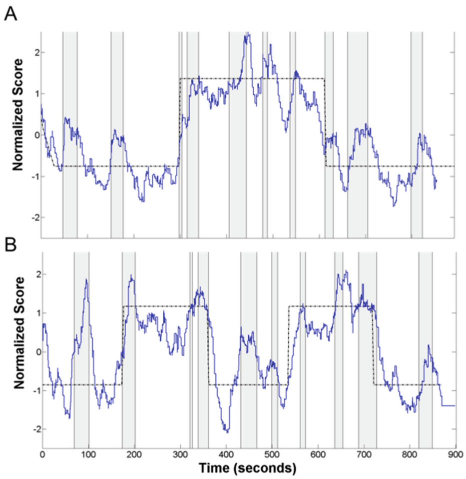

Fig. 2. Actual vs targeted score differences in the (A) low and (B) high challenge conditions. The square trace is the target scoring function and the irregular trace is a representation of the actual score differences observed. Shaded regions are times when the fog was enabled.

5 Initial Results

Though formal analyses are as yet incomplete, initial assessments are encouraging for the success of this paradigm. For instance, initial analyses on a subset (n = 7) of the subjects with processed ECG data revealed some expected patterns. In particular, heart rate variability, which has been shown to decrease with increased mental effort [10], was also shown to decrease when comparing baseline measures to recordings when performing the task; meanwhile, although to a much smaller effect, indications were present for a similar decrease in the fog as compared with periods when the fog was not present. Moreover, subsequent analyses on this same sample using K-means clustering suggested that those subjects who tended to increase the visual search area when going from low to high competitive challenge conditions, indicating greater visual searching for targets, also tended to show greater reductions in heart rate variability as the task difficulty increased. Finally, time frequency analyses comparing the low and high competitive challenge conditions on a separate subset of subjects (n = 9) with processed EEG are shown in Fig. 3. Here, a negative cluster (right-most scalp map; blue) was statistically significant ($p < 0.05$) while a positive cluster (left-most scalp map; red) was also observed although this did not reach significance for this small group size. In both clusters, the low-challenge condition appeared to be associated with greater event-related changes than in the high-challenge condition. Beta event-related synchronizations following an incorrect button-press response, as seen in these preliminary analyses, has been shown to increase in medial regions [11] and is likely reflected in the present data. One interpretation of such preliminary results is that the subjects were evaluating the meaning of their errors differently as a function of the perceived competitive challenge induced with this translational paradigm.

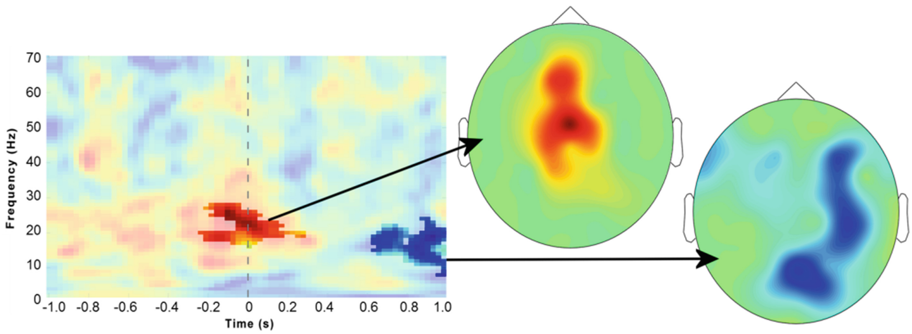

Fig. 3. Results from a time-frequency analysis of the EEG data recorded on the initial 9 subjects. Scalp map on the left shows positive event-related activation associated with the button press (vertical dashed line at 0 s on the time-frequency surface) and the scalp map on the right shows a negative event-related response following erroneous response trials.

6 Discussion

While the currently observed patterns in subject physiology and behavior have yet to be clearly and unequivocally associated with known affective state changes, these results reinforce the premise that performance differences can be modulated through an appropriately constructed competitive challenge. The results also suggest that subjects respond in different ways to the increased competition, as well as to actual changes in task difficulty (i.e., fog vs no fog). However, it is unlikely that any individual variable will embody the entire influences of competitive challenge. Rather a comprehensive view across multiple variables, simultaneously recorded and synchronized, should be employed. The paradigm described in this work was effective in reaching its initial objectives and well-suited for such operational analyses as suggested by the preliminary results.

Acknowledgements. The authors would like to our colleagues in the Translational Neuroscience Branch of the Army Research Laboratory for their help in designing and vetting this research project. This research was sponsored by the Army Research Laboratory and was accomplished under Cognition and Neuroergonomics Collaborative Technology Alliances (CaN CTA; Cooperative Agreement #W911NF-10-2-0022). The views and conclusions contained in this document are those of the authors and should not be interpreted as representing official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

1. Committee on Military and Intelligence Methodology for Emergent Neurophysiological and Cognitive/Neural Research in the Next Two Decades, N. R. C.: Emerging Cognitive Neuroscience and Related Technologies. The National Academies Press, Washington, DC (2008)
2. Committee on Opportunities in Neuroscience for Future Army Applications, N. R. C.: Opportunities in Neuroscience for Future Army Applications. The National Academies Press, Washington, DC (2009)
3. McDowell, K., Lin, C.-T., Oie, K.S., Jung, T.-P., Gordon, S., Whitaker, K.W., Li, S.-Y., Lu, S.-W., Hairston, W.D.: Real-world neuroimaging technologies. IEEE Access **1**, 131–149 (2013)
4. Kerick, S.E., McDowell, K.: Understanding brain, cognition, and behavior in complex dynamic environments. In: Schmorrow, D.D., Estabrooke, I.V., Grootjen, M. (eds.) FAC 2009. LNCS, vol. 5638, pp. 35–41. Springer, Heidelberg (2009)
5. Kerick, S., Metcalfe, J., Feng, T., Ries, A., McDowell, K.: Review of fatigue management technologies for enhanced military vehicle safety and performance. Technical report #ARL-TR-6571. US Army Research Laboratory, Aberdeen Proving Ground, MD (2013)
6. Rangel, A., Camerer, C., Montague, P.R.: A framework for studying the neurobiology of value-based decision-making. Nat. Rev. Neurosci. **9**, 545–556 (2008)

7. Ravaja, N., Saari, T., Turpeinen, M., Laarni, J., Salminen, M., Kivikangas, M.: Spatial presence and emotions during video game playing: does it matter with whom you play? Presence **15**(4), 381–392 (2006)
8. U.S Department of Defense Office of the Secretary of Defense. Code of federal regulations, protection of human subjects. 32 CFR 219. Government Printing Office, Washington, DC (1999)
9. U.S. Department of the Army. Use of volunteers as subjects of research. AR 70-25. Government Printing Office, Washington, DC (1990)
10. Allanson, J., Fairclough, S.H.: A research agenda for physiological computing. Interact. Compt. **16**(5), 857–878 (2004)
11. Koelewijn, T., van Schie, H.T., Bekkering, H., Oostenveld, R., Jensen, O.: Motor-cortical beta oscillations are modulated by correctness of observed action. Neuroimage **40**(2), 767–775 (2008)