Adapting the Taxon-Task-Taxon Methodology to Model the Impacts of Chemical Protective Gear

Shane T. Mueller
George Anno
Corey Fallon
Gene E. McClellan
Owen Price

Applied Research Associates, Inc.
1750 Commerce Center Blvd
Fairborn, OH 45324
smueller@ara.com

ABSTRACT: *The Taxon-Task-Taxon method (Anno et al., 1996) is a statistical modeling approach to predict performance decrements in response to various stressors. Our research is extending this approach to accommodate new more acute stressors associated with chemical protective gear, and new tasks with greater involvement of cognitive, perceptual, and motor function. In this paper, we describe the basics of the T3 method and our approach to adapting it, and give a illustrative example that shows how the method can be used to account for performance decrements associated with wearing protective gloves. This illustration provides a substantive way in which the current T3 method can be augmented to account for performance decrements in a new subdomain, but also provides lessons for extending the method to new stressors and performance domains.*

1. Background

Many cognitive and behavioral models aim to predict performance under new conditions, such as predicting performance for new tasks based on a measured set, or predicting performance on yet-to-be-built systems based on current performance, or predicting performance on a current task in response to new stressors. Our research program aims to understand the cognitive and behavioral performance decrements of chemical protective gear (i.e., Mission-Oriented Protective Posture; or MOPP) worn by U. S. warfighters in response to the threat or presence of chemical or biological agents. The intent of our models is to understand how new equipment may impact performance across a wide range of tasks to provide guidance for future suit design. Thus, we aim to predict performance decrements on a much wider range of tasks than can be effectively measured, under equipment conditions that have not yet been developed, and for novel combinations of new stressor.

1.1 Taxon-Task-Taxon (T3) Methodology

Our approach to simulating performance decrements in novel tasks under novel stress conditions is based on the Task-Taxon-Task methodology (T3; Anno, Dore, and Roth, 1996). The method works by assuming that performance degradation is mediated through a set of skill taxons (based on pioneering work by Fleishman,

1975). Any task is assumed to use these taxons to different extents, and each stressor is assumed to slow processes related to each taxon by different amounts. A predicted performance decrement for a particular stressor on a particular task can be computed by essentially computing the sum of the taxon-related decrements from the stressor, weighted by the relative importance of each taxon for the task. This statistical modeling approach is substantially less detailed than many agent-based modeling systems, but has advantages to the extent that it can be tied fairly closely to data, and that the effort for modeling new tasks or systems is fairly minimal (essentially a process of performing task analysis in order to develop ratings across skill taxa). This is important for our goal, because a single suit design will eventually be used across most branches and specialties of the U.S. military, and so a crude model that can predict across many tasks is preferred over a detailed model that can only predict a small range of tasks.

To use the method, a task Ti may be represented as a set of weights (e.g., between 0 and 5) relating to the relative importance over five taxa (attention, perception, physical, psychomotor, cognitive):

$Ti = [0,1,3,0,1]$

And similarly a stressor may be represented as a set of decrements across taxa (with 0 representing no impact, and values smaller than 0 representing the increase in log(RT1/RT0) ratio)

$$Sj = [-.05, -.01, -.2, -.05, -.1]$$

Here, Ti would represent a task with moderate physical requirements, and low requirements on other taxa. If Ti is assumed to take on unit of time, then the T3 model would assume that under stressor Sj, log(1/RT) of the task would be impacted by a factor of $(0(-.05) + 1(-.01) + 3(-.2) + 0(-.05) + 1(-.1)) = -.71$, which is a factor of 2.03. Thus, the large decrement high importance of the physical taxon, coupled with the large impact of the stressor on physical abilities would essentially double the time taken to perform the task.

The benefit of this method is that once careful assessment of the taxonomic weights are provided for a set of tasks, the impact of a particular stressor can be assessed using standard regression techniques (assuming a wide enough range of input tasks is available). Thus, the data fitting is a statistical process, although the decrements obtained could be used in other types of models. For example, along with its original use in predicting hypothesized impacts of chemical agents on soldier performance (e.g., Anno et al., 1996), this same method forms the basis for how the IMPRINT tool predicts performance decrements (Allender et al., 1997) for a number of stressors (MOPP, heat, cold, noise, and sleeplessness), although IMPRINT uses a set of nine taxons.

The T3 method was originally designed to predict behavioral decrements from toxic chemicals, based on a set of mediating symptomology. Such stressors have large-scale effects that may be well captured by global skill taxons. However, we are extending this method to account for the physical and especially cognitive stressors associated with chemical protective gear. Such stressors can have a much more acute impact on task performance. For example, one part of the MOPP suit is the gas mask and goggles, which have a well-understood impact on peripheral vision. Another component is butyl-rubber gloves, which impact a number of dexterous behaviors across specialties (see Mueller, et al., 2008a, 2008b). For such stressors, global taxons such as 'psychomotor' or 'perceptual' may no longer be sufficient to make useful predictions about performance decrements.

Along with the need to augment or change the current skill taxonomy, another problem for the T3 method is that as tasks become more complex and the stressors more acute, one may need better representations of tasks to make useful predictions about performance

decrements. Next, we will describe our approach to representing tasks.

2. Task-Goal-Operator-Taxon Analysis

One limitation of the original T3 method is it represents any task as a weighting across skill taxons. This may be appropriate for gross prediction of blunt stressors on highly constrained tasks, but it may be inappropriate for understanding the acute stressors of MOPP gear on detailed cognitive work. We have developed a task analysis method based on earlier GOMS methodologies (John & Kieras, 1994, Gray et al., 1993) by which we take a task and represent it as a critical path in a subgoal network (see Schweickert, Fisher, & Proctor, 2003) where each subgoal is accomplished by an operator, and each operator has a set of weights across relevant taxa (see Mueller et al., 2009a, for more detail). TGOT is similar to GOMS (Goal-Operator-Method-Selection rules) analysis in that is based on logical analysis of goals and subgoals which are traced to a set of operators. However, it differs because it uses a set of bottom-level operators that are tied to the task context, rather than low-level operators tied to an architecture. The point of TGOT analysis is to get to a level at which a task can be described in terms of its taxa, such that a stressor will have a linear impact on its time-to-perform. Thus, for GOMS, an operator is like a molecule: it can not be broken down further without changing its essence. For TGOT, an operator is like a mineral sample: any further subdivision will lead to identical parts in terms of the taxon distribution.

The use of a task network to represent tasks is important because of the ways in which we have hypothesized that protective gear may slow task performance. A partial list includes: First, the additional mass may simply make motor movement slower. Second, limited range-of-motion or perception may require taking new sets of actions (e.g., moving head to see in periphery). Third, reduced precision may lead to more errors which need to be corrected (e.g., mistaken key entry on keyboard). Fourth, wearing gear may place the wearer into a 'novice' performance mode as they grow accustomed to doing work under new conditions; eliminating automaticity gains. Fifth, gear may represent an attentional draw stemming from discomfort or additional self-monitoring required. Sixth, biophysical metabolic processes (heat, oxygen, bloodflow CO_2 maintenance, etc) may produce neurophysiological inefficiency or physical fatigue that impacts task performance. Seventh, the wearer intentionally and strategically slow down to avoid costly immediate error correction or long-term fatigue.

Although some of these sources may be well-captured by describing a high-level task as a set of operators, others are not. For example, intentional strategic

slowing may work to even out performance over a long period of time, rather than having fast performance initially and very slow performance later. So, one may observe slowing on a task in response to wearing MOPP gear, but the source of that slowing is strategic rather than physical. More critically, strategic shifts in task performance may also stem from limited mobility or limited sensory input. This type of shift may change the operators associated with performing a task, and may change the critical path in task performance. So, a stressor may not only change how long it takes to perform each step of a task, but it may also change the number of steps. An example of this in the context of manual dexterity will be show n in Section 3. Finally, stressors that impact accuracy may produce highly non-linear effects on certain aspects of a task, because slowing could stem primarily from error correction rather than slowed operation. Some type of task network analysis is necessary to understand whether that type of impact will have a large impact on overall task performance.

3. Example: Impact of Protective Gear on Human Dexterity

As an illustration, we will examine how the T3 method can be deployed to model human dexterity data. The original method included only one taxon (psychomotor) that can reasonably be used to describe performance in dexterity tasks. .Imprint incorporates two taxa (fine motor discrete and fine motor continuous), and assumes that only discrete action is impacted by protective gear. Such an example raises several questions. First, is a single taxon sufficient to capture the performance degradation on manual tasks associated with protective gear; and second, are there ways to know, *a priori* the extent to which a dexterity task will be impacted by a stressor?

As a first step, we present in Table 1 a set of proportional decrements for various motor dexterity tasks. In this Table, the performance decrement represents (Time with gloves)/(time in bare hands), so that a value of 1.0 would indicate no slowing from gloves, and larger values indicate larger impacts.

What can be said about the skill taxa necessary to capture these decrements? First, the one relevant taxon used previously (psychomotor) is probably insufficient. Certainly, one could assume that those tasks with greater decrements simply have higher psychomotor loadings. However, this is probably at odds with the ratings one would give a priori, and so is not very useful. For instance, it is probably unrealistic to say that those manual tasks which see little or no impact from protective gloves do not require psychomotor skill, and it would be difficult to predict a priori which types of tasks will have greater or lesser decrements,

especially when the decrements for similar tasks can vary so much.

Table 1: Performance decrements of various dexterity tasks.

Test	Perf. Decr.	Grasp	Touch	Pred
O'Connor Finger Test[12456]	1.14-1.72	5	1	1.29
Purdue Pegboard[126]	2.4-3.4	5	5	**1.6**
Minnesota Dexterity 1 hand[3]	1.17	2	3	1.27
Minnesota Dexterity-2 hand[4]	1.2-1.37	3	3	1.33
Manual Pursuit Rotor[1]	1.05	1	1	1.09
M16A1 Dis-Assembly[5]	1.24	3	3	1.33
M16A1 Assembly[5]	1.24	3	3	1.33
Find page in book[3]	1.25	3	3	1.33
1-5 number keypad entry[3]	1.09	1	1	1.1
Hunt-and-peck word typing[3]	1.22	1	3	1.23
Touch word-typing[3]	2.07	1	5	**1.37**
Typing response[3]	1.70	1	5	1.37
Mouse tracking[3]	1.15	1	3	1.23
Mouse—aimed movement[3]	1.01	1	1	1.1
Cord & Cylinder[2,4]	1.5-1.76	5	3	1.44
Bennet Dexterity test[4]	1.0-1.09	1	1	1.1
Pick up cylinder (20 mm+)[3]	1.05	1	1	1.1
Pick up cylinder (1 to 20 mm)[3]	1.25	3	3	1.3

[1]Bensel et al., 1987; [2]Taxiera et al., 1990; [3]Unpublished data by present authors; [4]McGinnis, Bensel, & Lockhart, 1973; [5]Garrett et al. 2006; [6]Johnson & Kobrick, 1997.

Note: Model fit excluded Purdue pegboard and touch-typing, which we assumed would have strategy shifts in response to protective gloves.

The two taxa used by IMPRINT are somewhat better, but they simply assume that 'continuous' tasks do not slowing, which could capture the small effects on the pursuit rotor and mouse aimed movement, but would miss the mouse tracking impact. As a first hypothesis, we propose that a way to capture these impacts would be to hypothesize two taxa: one related to grasping, and one related to the sense of touch. Initial ratings on the task for these taxa are provided in Table 1.

The Grasping taxon is important because picking up small objects has a moderate impact (25%) on performance, and this is a component that is present to in many of the tasks in Table 1. Loss of touch-sense could have a large impact depending on the context, because it may require costly error correction or strategy shifts. We hypothesize that this is partly responsible for the large decrements seen in typing (and indirectly, the Purdue task). Here, loss of touch sense is devastating. It can prevent touch-typing, which means that the errors one makes are not seen until it is very costly to correct. A typist must choose to either type, check for errors, and then correct errors, or slow down to a degree such that errors are not made (perhaps relying on visual and auditory feedback instead of touch sense). Either way, performance will slow substantially. The smallest impact seen on typing tasks was for number keypad entry: these were done hunt-and-peck style in both conditions, and the spacing of the number pad is big enough to avoid many mistakes. In essence, number-keypad entry would depend little on touch sense, whereas touch-typing relies heavily on it to know whether ones fingers are on the correct keys.

The Purdue test is interesting because it contains many of the same components measured in other tests, such as picking up small cylinders and placing them in holes or posts, which we showed to have a performance decrement of about only 25%. Yet the Purdue test had a substantial decrement at least ten times larger than these. What then can account for the difference? To answer this,, we need to understand better what the task involves.

The basic Purdue task involves four consecutive operations: 1. pick up and insert post; 2. pick up and insert washer; 3. pick up and insert sleeve; 4. pick up and insert second washer. Each consecutive step is performed by a different hand, so performance may be able to overlap substantially: Figure 1 illustrates how these four tasks may overlap because they use different hands.

Figure 1: Hypothesized subgoals to perform Purdue Pegboard task.
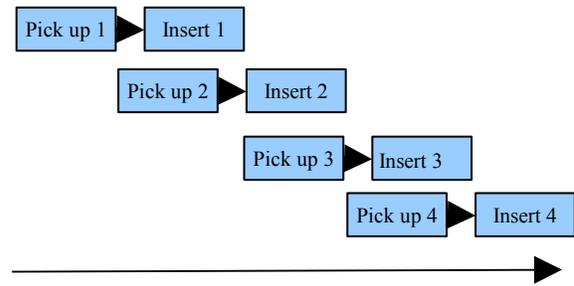
Total time to perform this task could be modeled as the sum (with p indicating pick-up time and *i* indicating insert time) of roughly $p1 + i1 + i2 + i3 + i4$.

However, for performance like this to occur, one needs to assume that these two tasks can be easily overlapped. Without protective gloves, the 'pick up' subtask might be thought of as performed by two operators, such as: move hand to tray; grasp object by feel. If we were to make a prediction about the performance decrement based on these operators using standard T3 methodology, we would find that overall task decrement should be driven by individual decrement for either the insert or pick up task (whichever requires more time). If we assume these operators have decrements of about 25%, the time to perform the overall sequence would increase by about 25%. This of course does not match the empirical finding that performance is slowed by a factor of 2 to 3.

However, task overlapping may not be possible with protective gloves, because limited sensory input will prevent the tasks from being overlapped. Thus, slowing in this task may stem from a shift to a non-overlapping performance strategy necessitated by reduced sensory impact. The sequence would be stretched out, as shown in Figure 2.
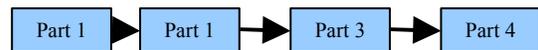

Figure 2: Hypothesized sequence of goals to perform Purdue Pegboard task with protective gloves.

Now, each pick up/insert subgoal must be achieved serially, and each of those subcomponents may slow as well. A reasonable estimate for the slowing would be that the task time would double, plush each component should increase by 25%, producing an estimated performance impact of 2.5, (instead of the 1.25 estimated from each individual operation).

To assess the extent to which the two dexterity taxa can account for performance decrements, we applied the T3 method as described by Anno et al. (1996). To estimate the impact I for each task, $\log(1/I)$ was computed, ensuring that all decrements would be negative numbers. Next, a linear regression model was fit to predict $\log(1/I)$ based on the two performance taxa ("grasp" and "touch"), excluding the Purdue and touch typing tasks because they were thought to involve strategy shifts. The intercept of the model was set to 0, as an intercept would simple amount to a generic decrement for all tasks. This regression was reliable ($F(2.14)=55$, $p<.01$) with an adjusted $R^2=.87$. The two predictors were reliable $p<.05$ (grasp= -0.04, $t(14)=-2.8$, $p=.01$; touch=-.054, $t(14)=3.9$, $p<.01$). These coefficient values indicate that each rating unit of the taxon reduces log-inverse-proportional performance by about .04-.05. Because for small values of p, $\exp(-p)$ approximates $1+p$, this means that each level of the rating scale slows performance by about 5%. Predicted performance values for each task are also printed in Table 1, along with the predictions for the two excluded task (shown in bold).

It should be noted that this method tends to underestimate the impact of those stressors with large decrements. The performance model described has a limited upper level, with log-inverse-proportion having a maximum decrement of about .45 (or 1.6). Most likely, to accommodate larger impacts, one must incorporate simple notions of strategy shifts (such as we argued for in the Purdue task), or costly error-recovery processes that are outside the linear model used in the T3 process. As a rough guide, in order to predicted a decrement of 3.0, the Purdue task would need a touch value of about 22, which is well beyond the end of our scale.

4. Discussion
The T3 method offers a simple statistical method for predicting coarse decrements across tasks in response to a number of stressors. Although predictions needing finer precision may require agent-based modeling with systems such as EPIC (e.g., Meyer et al., 2001, in the context of age-related stressors), we are developing ways to adapt the process to enable prediction for acute stressors related to MOPP gear, and involved with more perceptual, motor, and cognitive tasks. These adaptations take two forms. First, we are beginning to hypothesize new performance taxa that can be used to understand whether some task will see large decrements from protective gear. Second, we hypothesize that a more detailed task representation needs to be used, which can at least help identify whether a stressor will induce strategic shifts or costly error recovery processes.

We illustrated how these additional factors are important for extending the T3 method to the relatively simple domain of manual dexterity. In future and ongoing work, we are extending the method to tasks with stronger cognitive and perceptual components, which we believe will require similar additions to the T3 process.

5. References
Allender, Salvi, & Promisel (1997). Evaluation of Human Performance under Diverse Conditions.

Anno, G. H., Dore, M. A., and Roth, T. J. (1996). Taxonomic model for performance degradation in combat tasks. (DNA-TR-95-115) Alexandria, VA: Defense Nuclear Agency.

Bensel, C. K., Teixeira, R. A., & Kaplan, D. B. (1987). The Effects of US Army Chemical Protective Clothing on Speech Intelligibility, Visual Field, Body Mobility, and Psychomotor Coordination of Men (TR-87-037). Natick, MA: United States Army Natick Research. Development, and Engineering Center.

Fleishman, E. A. (1975). Toward a Taxonomy of Human Performance. American Psychologist, 30(12), 1127-1149.

Garrett, L., Jarboe, N., Patton, D. J., & Mullins, L. L. (2006). The Effects of Encapsulation on Dismounted Warrior Performance. Retrieved from http://stinet.dtic.mil/oai/oai? &verb=getRecord&metadataPrefix=html&identifier=ADA448610

Gray, W. D., John, B. E., & Atwood, M. E. (1993). Project Ernestine: Validating a GOMS analysis for predicting and explaining real-world task performance. Human-Computer Interaction, 8(3), 237-309.

John, B. E., & Kieras, D. E. (1994). The GOMS family of analysis techniques: Tools for design and evaluation. Rappot technique CMUCHI-94, 106.

Johnson, R., & Kobrick, J. (1997). Effects of Wearing Chemical Protective Clothing on Rifle Marksmanship and on Sensory and Psychomotor Tasks. Military Psychology, 9(4), 301-314. doi: 10.1207/s15327876mp0904_3

McGinnis, J. S., Bensel, C. K., & Lockhar, J. M. (1973). Dexterity afforded by CB protective gloves. US Army Natick Laboratories, Natick, MA, Report.

Meyer, D. E., Glass, J. M., Mueller, S. T., Seymour, T. L., & Kieras, D. E. (2001). Executive-process interactive control: A unified computational theory for answering 20 questions (and more) about cognitive ageing. European Journal of Cognitive Psychology, 13(1), 123-164.

Mueller, S. T., Zimmerman, L. A., McClellan, G. E., Crary, D. J., & Cheng, K. (2009). Cognitive Performance Prediction with the T3 Methodology,.

Mueller S. T., Zimmerman, L. Cheng, K., Crary, D., Simpkins, B., Gabbard, S., and McClellan, G. (2009). Establishing a set of cognitive tests to predict the effect of novel protection suit design concepts. Chemical Biological Defense (CBD) Science and Technology Conference, Dallas, TX, November 2009.

Schweickert, R., Fisher, D. L., & Proctor, R. W. (2003). Steps toward building mathematical and computer models from cognitive task analyses. Human factors, 45(1), 77.

Teixeira, R. A., & Bensel, C. K. (1990). The effects of chemical protective gloves and glove liners on manual dexterity (U). US Army Natick Research, Development and Engineering Center, Natick, Massachusetts, NATICK/TR/91/002.

Author Biographies

DR. SHANE MUELLER is Senior Research Scientist at Applied Research Associates in Fairborn, OH. He specializes in measuring and modeling human behavior, with emphasis in decision making, human performance, and memory. He is the developer of the PEBL test battery for measuring psychology performance.

GEORGE ANNO is an independent consultant on this research effort. He is the originator of the T3 method, and specializes in modeling the impacts of chemical and biological stressors on human performance.

COREY FALLON is a Staff Research Scientist at Applied Research Associates in Fairborn, OH. He specializes in applied human factors, with an emphasis on understanding the impact of new technology on individual and team workflow.

GEORGE ANNO is an independent consultant on this research effort. He is the originator of the T3 method, and specializes in modeling the impacts of chemical and biological stressors on human performance.

DR. GENE E. MCCLELLAN is Principal Research Scientist at Applied Research Associates in Arlington, VA. He is Director of the Health Effects/Medical Response group, and has led efforts for estimating battle casualties from NBC attacks in support of U.S. and NATO medical defense planners, and was Program Manager for the development of the medical NBC Casualty and Resource Estimation Support Tool (NBC CREST) for medical planning.

OWEN PRICE is Senior Research Scientist at Applied Research Associates in Raleigh, NC, where he develops the Multiple-Path Particle Dosimetry (MPPD) model, and assists with modeling and software development for other projects in the Health Effects and Medical Response group.