ARL-TN-1300 ● MAR 2026

Theory-Driven Big Data (TDBD) Approaches: Final Report

by Stephen R. Mitroff, Dwight J. Kravitz,
Chloe Callahan-Flintoft, Sean M. Fitzhugh, and Kelvin S. Oie

NOTICES

Disclaimers

The findings in this report are not to be construed as an official Department of the Army position unless so designated by other authorized documents.

Citation of manufacturer's or trade names does not constitute an official endorsement or approval of the use thereof.

Destroy this report when it is no longer needed. Do not return it to the originator.

ARL-TN-1300 ● MAR 2026

Theory-Driven Big Data (TDBD) Approaches: Final Report

Stephen R. Mitroff
The George Washington University

Dwight J. Kravitz
The George Washington University and U.S. National Science Foundation

Chloe Callahan-Flintoft, Sean M. Fitzhugh, and Kelvin S. Oie
DEVCOM Army Research Laboratory

REPORT DOCUMENTATION PAGE

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED	
		START DATE	END DATE
March 2026	Technical Note	16 July 2024	15 July 2025

4. TITLE AND SUBTITLE
Theory-Driven Big Data (TDBD) Approaches: Final Report

5a. CONTRACT NUMBER	5b. GRANT NUMBER	5c. PROGRAM ELEMENT NUMBER
W911NF-24-2-0188		
5d. PROJECT NUMBER	**5e. TASK NUMBER**	**5f. WORK UNIT NUMBER**

6. AUTHOR(S)
Stephen R. Mitroff, Dwight J. Kravitz, Chloe Callahan-Flintoft, Sean M. Fitzhugh, and Kelvin S. Oie

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
DEVCOM Army Research Laboratory
ATTN: FCDD-RLA-FA
Aberdeen Proving Ground, MD 21005

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TN-1300

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
DISTRIBUTION STATEMENT A. Approved for public release: distribution unlimited.

13. SUPPLEMENTARY NOTES
ORCIDs: Stephen R. Mitroff 0000-0003-3524-6396; Dwight J. Kravitz 0000-0002-8108-1329; Chloe Callahan-Flintoft 0000-0002-1874-2032; Sean M. Fitzhugh, 0000-0002-6283-2895; Kelvin S. Oie 0000-0002-0311-800X
This report represents the final report for cooperative agreement W911NF-24-2-0188 between ARL and The George Washington University.

14. ABSTRACT
Traditional theory-based methods are often reductive, struggling to detect smaller but critical effects, and their predictions frequently fail outside controlled laboratory settings. Conversely, while the statistical power of "Big Data" is immense, purely data-driven approaches can result in poorly generalizing "black box" models that limit explanatory interpretation and practical understanding. This project explored a novel perspective on human behavior—the Theory-Driven Big Data approach—by leveraging the complementary strengths of traditional theory-driven and contemporary data-driven research. Project efforts focused on addressing fundamental questions about human behavior to inform performance improvements in demanding situations, such as those faced by U.S. Soldiers. We applied novel, theory-driven analyses to a unique dataset of nearly 4 billion trials from over 15 million individuals playing the mobile game Airport Scanner, which provided an unprecedented opportunity to quantify the effects of established theoretical mechanisms, such as memory decay and task interference, and to examine the impact of standard experimental design practices like counterbalancing and randomization in a real-world context.

15. SUBJECT TERMS
individual differences, inter-individual variability, intra-individual variability, visual search, cognition, motor behavior, experimental methods, Humans in Complex Systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES
a. REPORT	b. ABSTRACT	C. THIS PAGE		
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UU	18

19a. NAME OF RESPONSIBLE PERSON	19b. PHONE NUMBER *(Include area code)*
Kelvin S. Oie	(520) 691-5705

STANDARD FORM 298 (REV. 5/2020)
Prescribed by ANSI Std. Z39.18

Contents

1. Introduction and Proposed Approach

The primary goal of this cooperative agreement was to explore a radical and novel perspective on human behavior—the Theory-Driven Big Data (TDBD) approach—by leveraging the collective strengths and resources from two complementary teams. Traditional theory-driven research approaches have yielded many important insights into human behavior and capabilities and hold substantial promise for real-world application, but their utility also has clear pitfalls. First, by taking highly reductive perspectives, traditional approaches and experimental methods in the human sciences have been used successfully to detect large effects (e.g., large differences in group or condition means), but struggle with smaller—often critical—effects. Second, potentially because they ignore the multiple factors that often drive these smaller but critical effects, their predictions frequently fail when applied to real-world problems outside of highly controlled laboratory conditions. Finally, they tend to focus on whether there is *any* effect, which is rendered moot by the power inherent in very large data—in which most effects will prove statistically significant.

The increased statistical power afforded by "Big Data" also constrains more recent data-driven approaches. While they can reveal a host of significant differences, the algorithmic methods used to optimize model performance result in poor generalization when the model is applied to different data or contexts. Moreover, many of these are "black box" models, with very large numbers of parameters that make understanding their contributions to model predictions and behavior practically impossible, thus limiting explanatory interpretation.

The proposition of this project is that leveraging the strengths of both traditional and Big Data approaches in tandem provides an opportunity to assess highly multidimensional real-time human behaviors in a way that outperforms the predictiveness, generalizability, and/or understandability of models produced through either theory- or data-driven approaches alone.

Thus, research efforts within this project aimed to address fundamental questions on human behavior that offered insights for improving human performance in demanding situations, like those faced by U.S. Soldiers. Projects were focused on exploring approaches that identified, implemented, and applied novel theory-driven approaches and analyses to large sample size data. Analytical efforts executed within this project leveraged both deep knowledge of human visual search behavior and a unique dataset comprising nearly 4 billion trials performed by over 15 million

individuals in a visual search-based, mobile gaming application, *Airport Scanner*.[*] The massive size of the *Airport Scanner* dataset afforded opportunities to explore TDBD questions that would be intractable—if not actually impossible—to ask in traditional laboratory settings. In this project, we pursued two research tasks aimed at quantifying the effects of both established theoretical mechanisms (e.g., memory decay and task interference) and standard experimental design practices (e.g., counterbalancing and randomization).

1.1 Task 1: Understanding the Joint and Unique Contribution of Decay and Interference

There is a long-standing debate about whether forgetting is best understood as a process of 1) decay, wherein the passage of time alone causes representations to become less stable; or 2) interference, wherein intervening events impinge on preexisting representations causing them to become less stable. There is evidence for both mechanisms, but almost never sufficient power and variability to investigate both simultaneously. This task uses the unique characteristics of the *Airport Scanner* dataset to probe this critical aspect of human behavior with unprecedented power and fidelity to yield insights into the underlying mechanisms.

1.2 Task 2: Optimizing Experimental Design

In most standard experiments, the pragmatic constraints on data collection (e.g., length of experiments and number of participants) limit the capability of the research team to balance possible factors that could impact the strength of the experimental design. It is well known that behavior is nonstationary and varies over time and experience; however, these factors are rarely controlled, let alone tested. Rather, components of the task (e.g., trial order and assignment of stimuli to conditions) are determined randomly, causing the variation in behavior to be smeared across conditions. This task uses the sheer volume of data and variation of individual participant experiences in the *Airport Scanner* dataset to quantify the benefit of counterbalancing and the cost of not doing it. This task will highlight the unique complexities involved in studying human behavior and yield insights into the proper approach to experimental design.

[*] (Kedlin Company; Bellevue, Washington) https://www.airportscannergame.com/

2. Research Accomplishments

Over the course of this award's period of performance, the George Washington University (GWU) and U.S. Army Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) research teams worked together to address goals for Task 1 and Task 2. This resulted in multiple accomplishments, including successful completion of all project deliverables and scientific advances, which have been reported in numerous knowledge products. These are itemized in the Appendix. At a high level, this award contributed to 5 manuscripts published or in-process, and 18 professional talks and presentations at scientific conferences or other venues. Fourteen of these works were jointly authored by GWU and ARL collaborators. Ongoing efforts that extend from the research results and advances made within this project will further expand these numbers.

2.1 Deliverables and Notable Advances

2.1.1 Task 1: Joint Effects of Decay and Interference

We directly investigated questions about the joint and unique contribution of decay and interference on task performance using data from the *Airport Scanne*r visual search task in several projects. In this task, players take in the role of airport security screeners and repeatedly search bags for prohibited items. Because of its size, this dataset afforded the necessary power to simultaneously examine the impact of delay and interfering events in a statistically rigorous way, despite the complexity of the task and the variation observed across the behavior of individual players. Projects funded under this award extended foundational ideas explored under previous funding (e.g., Callahan-Flintoft et al. [2023] and Cox et al. [2023]), with several projects currently being prepared for submission and forming the core scientific contributions of the upcoming doctoral thesis of one of the research team's graduate students (Emma Siritzky) who was supported by this award.

Ms. Siritzky's project examines the hysteresis effects that arise from both task-relevant (target identity) and task-irrelevant (bag) elements of the visual search and their interaction. In the first experiment (Siritzky et al. 2024; Siritzky 2025), we examined the impact of the elements of a trial on the search performance in the trial that immediately followed. We found an impact of both the target and bag, with faster response times (RTs) when either of these elements matched between the two trials than when they did not. Critically, the benefit when both elements matched exceeded the summed benefit of the individual elements, indicating associative coding. In subsequent experiments (Siritzky 2025) we showed that the benefit of matching targets (task-relevant information) was more resistant to both delay (i.e.,

both the number and content of intervening trials and the time elapsed between exposure to matched targets) and interference effects (i.e., exposure to matching or non-matching targets and bags on intervening trials). This resilience was further strengthened when both task-relevant and task-irrelevant elements matched, which is consistent with associative coding conveying a benefit over and above the individual elements. We are in the process of running the final analysis that will directly compare delay and interference. The project highlights how the impact of either factor is determined by the strength of the representation, with items that drive behavior and associations with those items being more resilient.

2.1.2 Task 2: Optimizing Experimental Design

Across all projects conducted as part of this cooperative agreement, a core theme has been a focus on optimizing experimental design (e.g., Kravitz and Mitroff [2023] and Siritzky et al. [2023]). Leveraging the large dataset afforded by the *Airport Scanner* game, it has been possible to 1) explore the impact of different experimental design choices and 2) demonstrate how optimizing design parameters can control confounds and increase clarity. One example comes from a project with a journal article currently under review at *Attention, Perception, & Psychophysics* (Malykke et al. 2026).

Errors are a fundamental aspect of human performance, and it is vital to understand their causes so they can be reduced or addressed. Critically, errors rarely occur in isolation but rather within behavioral sequences, which are known to shape subsequent behavior. However, understanding the full time-course of the impact of errors has been challenging given data limitations. This project compared two groups of precisely matched participants ($n > 35{,}000$): those who made an error versus those who did not make an error in a short object discrimination task. The matching among groups was extremely careful so that all aspects were controlled (e.g., timing, stimulus presentations, and participant counts by scenario), and linear mixed-effects models revealed error-induced effects both preceding and following errors. First, the error group showed pre-error speeding up to eight trials before the error occurred. Second, post-error slowing occurred and rapidly diminished over trials. Finally, leave-one-out cross-validation models using individual trial behavior predicted an upcoming error eight trials before the error occurred. This project highlights the systematic nature of behavioral changes around errors and demonstrates the power of optimized experimental designs.

2.1.3 Theory-Informed Analyses of Individual Differences

Previous analyses of very large ($N > 15$ million) datasets were extended upon to explore new approaches to evaluating theory-informed models that capture the

effects of multiple factors, which simultaneously affect complex human task behavior. This decision was made due to the increased data requirements for models that include multiple variables. Comparisons were made of the fits and parameters of multifactorial models describing dimensions of task context, participant experience, and individual differences. Initial results demonstrated that all tested variables contributed meaningfully to observed performance variation, consistent with the project's working hypothesis that human behavior is simultaneously shaped by multiple factors. Importantly, these results also showed the critical advantage of using very large datasets in analyses that reflect the high variability and multidimensionality of real-world human behavior (Siqi-Liu et al. 2024). These results provide a foundation for theory-driven approaches to develop models that could improve the current state-of-the-art approaches for the prediction of individual performance and long-term ability.

3. References

Callahan-Flintoft C et al. Characterizing exposure history effects on visual search performance using big data. OPAM 31: Object Perception, Attention, and Memory Workshop; 2023 Nov 16; San Francisco, CA.

Cox PH et al. Characterizing exposure history effects on visual search performance using big data. OPAM 31: Object Perception, Attention, and Memory Workshop; 2023 Nov 16; San Francisco, CA.

Kravitz DJ, Mitroff SR. Quantifying, and correcting for, the impact of questionable research practices on false discovery rates in psychological science. J Reprod Neurosci. 2023. https://doi.org/10.36850/jrn.2023.e44

Malykke SB, Siqi-Liu A, Kravitz DJ, Mitroff SR. Predicting errors before they occur using trial-by-trial dynamics of pre- and post-error behavior. Atten Percept Psychophys. Forthcoming 2026.

Siqi-Liu A et al. Comparing apples to oranges to bananas: a big data approach to understanding the joint influences of stimulus properties, trial history, and individual differences. VSS 2024: 24th Annual Meeting of the Vision Sciences Society; 2024 May 17–22; St Pete Beach, FL.

Siritzky EM. The impact of decay and interference on incidental associative learning for task-relevant and task-irrelevant information in visual search. The George Washington University (US), Department of Psychological & Brain Sciences, Cognitive Neuroscience Brownbag Series; 2025.

Siritzky EM et al. Incidental learning of task-relevant and task-irrelevant information in visual search reveals nuanced effects of decay and associative processing. OPAM 32: Object Perception, Attention, and Memory Workshop; 2024 Nov 21; New York, NY.

Siritzky EM et al. Standard experimental paradigm designs and data exclusion practices in cognitive psychology can inadvertently introduce systematic "shadow" biases in participant samples. Cog Res: Prin Implications. 2023 Oct 21;8(1):66.

Appendix. Project Knowledge Product Metrics

A.1 Research Team

The George Washington University (GWU)

Co-Principal Investigator:	Stephen R. Mitroff
Co-Principal Investigator:	Dwight J. Kravitz
Postdoctoral Fellow:	Audrey Siqi-Liu
Ph.D. Graduate Student:	Samoni Nag
Ph.D. Graduate Student:	Emma Siritzky
Ph.D. Graduate Student:	Justin Grady
Ph.D. Graduate Student:	Sarah Malykke
Undergraduate Researcher:	Sarah Hodgdon
Post-baccalaureate Researcher:	Victor Platzer

DEVCOM Army Research Laboratory

Co-Principal Investigator:	Kelvin S. Oie
Co-Investigator:	Chloe Callahan-Flintoft
Co-Investigator:	Sean M. Fitzhugh
Army Educational Outreach Program Postdoctoral Fellow:	Ellie Robbins
Oak Ridge Associated Universities Postdoctoral Fellow:	Melanie Bancilhon

A.2 Accomplishments by Category

Publications and Manuscripts

Ritchie JB, Wardle S, Vaziri-Pashkam M, Kravitz DJ, Baker CI. Rethinking category-selectivity in human visual cortex. Cog Neurosci. 2025:1–28. https://doi: 10.1080/17588928.2025.2543890

Kravitz DJ, Mitroff SR, Callahan-Flintoft C, Oie KS. Crowdsourced data collection opens new avenues for the behavioral sciences to impact real-world applications. Policy Insights Behav Brain Sci. 2024;11(2):123–131. https://doi.org/10.1177/23727322241274745

Callahan-Flintoft C et al. Repeated exposure to task-relevant and task-irrelevant information—and their interaction—affect visual search performance. Manuscript in press at Atten Percept Psychophys.

Fitzhugh SM et al. Leveraging outliers to predict rare behaviors in a massive sample. Manuscript under revision.

Malykke SB, Siqi-Liu A, Kravitz DJ, Mitroff SR. Predicting errors before they occur using trial-by-trial dynamics of pre- and post-error behavior. Manuscript under review at Atten Percept Psychophys.

Conference Podium Presentations

Siqi-Liu A et al. A big data approach to modeling the joint influences of stimulus, trial hysteresis, and individual differences on variability in trial-by-trial visual search performance. Psychonomic Society 65th Annual Meeting; 2024 Nov 21–24; New York, NY.

Siritzky EM et al. Incidental learning of task-relevant and task-irrelevant information in visual search reveals nuanced effects of decay and associative processing. OPAM 32: Object Perception, Attention, and Memory Workshop; 2024 Nov 21; New York, NY.

Conference and Workshop Poster Presentations

Dochnal M, Siqi-Liu A, Malykke SB, Mitroff SR. Using mouse tracking data to examine adaptation of visual search behaviors to transitions in task complexity. GWU CCAS Research Showcase; 2025 Apr 2; Washington, DC.

Grady JN, Siritzky EM, Kravitz DJ, Mitroff SR. Characterizing the independent and joint impacts of previous category evidence and visual and semantic similarity on visual attention. VSS 2025: 25th Annual Meeting of the Vision Sciences Society; 2025 May 16–20; St Pete Beach, FL.

Malykke SB, Siqi-Liu A, Oie KS, Kravitz DJ, Mitroff SR. The time course of an error: benefits of examining pre-error speeding and post-error slowing simultaneously. VSS 2025: 25th Annual Meeting of the Vision Sciences Society; 2025 May 16–20; St Pete Beach, FL.

Callahan-Flintoft C, Cox PH, Kramer MR, Mitroff SR, Kravitz DJ. A high-fidelity quantification of the time course of learning consolidation from a massive visual search dataset. HCxS Human-Autonomy Teaming VIP Day; 2024 Aug 24; Aberdeen Proving Ground, MD.

Carter EC et al. Theory-driven big data for Soldier characterization. HCxS Human-Autonomy Teaming VIP Day; 2024 Aug 24; Aberdeen Proving Ground, MD.

Cox PH, Callahan-Flintoft C, Kramer, MR, Mitroff SR, Kravitz DJ. A high-fidelity quantification of the time course of learning consolidation from a massive visual search dataset. VSS 2024: 24th Annual Meeting of the Vision Sciences Society; 2024 May 17–22; St Pete Beach, FL.

Malykke SB et al. Changes in response times before and after making a mistake reveal distinct motor and cognitive mechanisms of error monitoring. OPAM 32: Object Perception, Attention, and Memory Workshop; 2024 Nov 21; New York, NY.

Robbins E, Kravitz DJ, Hereth A, Oie KS, Mitroff SR. PARADIGMS: a new system for precise, accurate, and efficient measurement of human behavior: collaboration opportunities. University of Colorado Human-Machine Teaming Paradigms Meeting; 2024 Mar 15–16; Boulder, CO.

Siritzky EM, Siqi-Liu A, Kravitz DJ, Vickery TJ, Mitroff SR. Examining differential effects of target and context repetition in visual search: insights from a big data approach. VSS 2024: 24th Annual Meeting of the Vision Sciences Society; 2024 May 17–22; St Pete Beach, FL.

Siqi-Liu A et al. Comparing apples to oranges to bananas: A big data approach to understanding the joint influences of stimulus properties, trial history, and individual differences. VSS 2024: 24th Annual Meeting of the Vision Sciences Society; 2024 May 17–22; St Pete Beach, FL.

Siqi-Liu A et al. A big data approach to understanding the joint influences of stimulus properties, trial history, and individual differences [Presented by Oie KS]. HCxS Human-Autonomy Teaming VIP Day; 2024 Aug 24; Aberdeen Proving Ground, MD.

Other Invited Talks

Mitroff SR. Using "theory-driven big data" to advance psychology research. National Academies, Board on Behavioral, Cognitive, and Sensory Sciences. National Academies of Science and Engineering; 2025.

Brownbag Talks

Grady JN. Revealing precise and intricate influences of prior history on visual attention. The George Washington University (US), Department of Psychological & Brain Sciences, Cognitive Neuroscience Brownbag Series; 2025.

Siritzky EM. The impact of decay and interference on incidental associative learning for task-relevant and task-irrelevant information in visual search. The George Washington University (US), Department of Psychological & Brain Sciences, Cognitive Neuroscience Brownbag Series; 2025.

Callahan-Flintoft C. Understanding attention in more realistic, mission-relevant scenarios with a computational cognitive model. The George Washington

University (US), Department of Psychological & Brain Sciences, Cognitive Neuroscience Brownbag Series; 2024.

Malykke SB. Mapping the time course of errors in task performance. The George Washington University (US), Department of Psychological & Brain Sciences, Cognitive Neuroscience Brownbag Series; 2024.

Awards

Siritzky EM. Best talk award: object perception, attention, and memory workshop. OPAM 32: Object Perception, Attention, and Memory Workshop; 2024 Nov 21; New York, NY. (Award = $150)

Degrees and Promotions

- Hodgdon SV. Accepted to Lehigh University Ph.D. program, June 2025
- Siritzky EM. Comprehensive examinations, June 2025
- Malykke SB. Comprehensive examinations, Apr 2025
- Malykke SB. Master's degree, Dec 2024
- Siritzky EM. Master's degree, Dec 2024

List of Symbols, Abbreviations, and Acronyms

ARL	Army Research Laboratory
DEVCOM	U.S. Army Combat Capabilities Development Command
GWU	The George Washington University
RT	Response Time
TDBD	Theory-Driven Big Data

1 (PDF)	DEFENSE TECHNICAL INFORMATION CTR DTIC OCA
1 (PDF)	DEVCOM ARL FCDD RLB CI TECH LIB