ARI-SS-SR 2024-09
Scientist Series
Scientific Report

April 2024

Active Inference: A Competency for Making Decisions in Uncertain Situations

James M. Nye
Cary Stothart
Rhett Graves
U.S. Army Research Institute

Alex Francisco
Jared Peterson
Consortium of Universities of Washington

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

SCOTT SHADRICK, Ph.D.
Acting Director

Research accomplished for the Department of the Army by:

U.S. Army Research Institute for the Behavioral and Social Sciences,
Fort Leavenworth Research Unit

Research Unit Chief
Dr. Rhett Graves (Fort Leavenworth Research Unit)

Team Leader
Dr. Kingsley C. Ejiogu

Technical Reviewers
Dr. Rachel Amey, U.S. Army Research Institute
Dr. Lee Bedford, Center for Army Leadership

DISPOSITION

This Technical Report has been submitted to
the Defense Technical Information Center (DTIC)

REPORT DOCUMENTATION PAGE

1. REPORT DATE (Month Year) March 2024	**2. REPORT TYPE** Scientific Report, Final	**3. DATES COVERED (Month Year)** START DATE September 2018	END DATE December 2021

4. TITLE AND SUBTITLE
Active Inference: A Competency for Making Decisions in Uncertain Situations

5a. CONTRACT NUMBER W911NF-21-C-0040	**5b. GRANT NUMBER**	**5c. COOPERATIVE AGREEMENT NUMBER**

5d. PROGRAM ELEMENT NUMBER 622785	**5e. PROJECT NUMBER** A790	**5f. TASK NUMBER** 0901	**5g. WORK UNIT NUMBER**

6. AUTHOR(S)
Nye, James, M; Stothart, Cary; Graves, Rhett; Francisco, Alex; Peterson, Jared

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, VA 22060-5610	**8. PERFORMING ORGANIZATION REPORT NUMBER** SS-SR 2024-09

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Research Institute for the Behavioral and Social Sciences 6000 6th Street (Bldg. 1464 / Mail Stop: 5610) Fort Belvoir, VA 22060-5610	**10. SPONSOR/MONITOR'S ACRONYM(S)** ARI	**11. SPONSOR/MONITOR'S REPORT NUMBER(S)** SS-SR 2024-09

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Cary Stothart, Fort Leavenworth Research Unit

14. ABSTRACT
Army officers need to make decisions while facing unprecedented levels of volatility, uncertainty, complexity, and ambiguity. We propose that active inference—the ability to infer the true state of the world by revising beliefs that generate invalid predictions—is pertinent to Army officers' ability to make sense of evolving situations they have never encountered before and about which they have incomplete information. A prototype active inference measure—the Decisions Over Time (DOT) task—was developed to measure how participants learn a complex structure of rules based solely on iterative exposures to stimuli that change systematically over time. Participants inferred the rules governing changes in stimuli across multiple iterations by making choices and inferring from outcomes, developing a mental model of the rules governing the pattern of change. We found that the DOT task differentiates active inference ability levels along two distinct dimensions: (a) an ability to learn from prediction errors, and (b) a sensitivity to prediction error. Future research will further refine the efficacy of the DOT task for assessing Army officers' capability to make sense of evolving and uncertain situations. The DOT task may be applied in assessing officers' active inference preferences and strategies to inform tailoring of decision making training to individual learning needs.

15. SUBJECT TERMS
Strategic Thinking, Action Learning, Comprehensive Information Gathering, Thinking-in-Time, Leadership Requirements Model, Officer Assessment

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**
a. REPORT Unclassified	**b. ABSTRACT** Unclassified	**C. THIS PAGE** Unclassified	Unlimited Unclassified	48

19a. NAME OF RESPONSIBLE PERSON Dr. Rhett Graves	**19b. PHONE NUMBER (Include area code)** 913-547-6811

ACKNOWLEDGEMENT

We would like to acknowledge Dr. Mark Poirier for his comments and suggestions on this project. We would also like to broadly acknowledge all the Army personnel and ARI staff who make all our research possible. Finally, we thank the many Soldiers who participated in this research.

ACTIVE INFERENCE: A COMPETENCY FOR MAKING DECISIONS IN UNCERTAIN SITUATIONS

EXECUTIVE SUMMARY

Research Requirement:

Army officers make decisions within operational environments composed of complex interdependencies that are evolving over time. Given this situation, officers must make decisions while relying on information about their environment that is often incomplete. To be effective decision-makers, officers develop and revise their situational understanding from moment to moment, inferring patterns from past and present events to anticipate probable future events. Working with incomplete information, officers always perceive, decide, and act under some degree of uncertainty. Moreover, officers are themselves causal agents, contributing to the complexity and ambiguity of the environment in which they are acting. Every action they take has consequences that serve to further shape their future options. Their understanding of the operational environment is changed every time they decide, act, and then observe the consequences of their actions. When officers observe the consequences of their actions, it serves to shape their mental model of the environment. This revised mental model informs what they will do next.

We propose that active inference—the ability to infer the true state of the world by revising beliefs that generate invalid predictions—is pertinent to Army officers' capability to make sense of novel, complex, and evolving situations. From a cognitive neuroscience perspective, active inference describes a mechanism by which the mind seeks to resolve uncertainty and minimize surprise (i.e., if an outcome is highly surprising, then one's beliefs might be generating invalid predictions and should therefore be revised). Active inference is an ongoing process of iteratively predicting future outcomes based on our current understanding of a situation, our actions in that situation, identifying errors in our predictions about the outcomes of our actions, and using those errors to further calibrate the existing belief which had informed our initial prediction. By developing a measure of active inference capability in Army officers, we provide a means by which officers' competencies supporting decision-making under uncertainty may be assessed and further enhanced.

Approach:

Here, we present and evaluate a novel active inference measure: the Decisions Over Time (DOT) task. The DOT task requires participants to learn a complex structure of rules governing the changing patterns of stimuli presented. Participants must construct in real time a mental model of the rules governing how stimuli are being presented, learning through trial-and-error to predict outcomes more accurately based on recognizing their previous mistakes and updating their assumptions. By creating and validating a task to assess officers' ability to revise their beliefs based on performance feedback, the Army may be better able to assess Army leaders' facility in adapting to new situations in which prior experience and training have ceased to provide a reliable basis for prediction.

Findings:

The DOT task was able to differentiate levels of ability in task performance as well as two distinct methods of effective problem-solving. Four types of performers were identified, namely adapters, overlearners, satisficers, and guessers. Adapters developed a useful predictive model early on and then updated it when their actions elicited errors, as evidenced by their learning one rule, and then adapting to new rules. Overlearners also developed a useful predictive model, but then were challenged to update their model as task requirements changed, as evidenced by their learning the first rule but failing to adapt when it changed. Satisficers quickly developed a good-enough, albeit imperfect model, as evidenced by identifying a single response throughout the task that happened to be correct 66.6% of the time. Finally, guessers never developed a useful model, as evidenced by their failing to learn any rules. Overall, our results suggest that the DOT task might be able to differentiate active inference ability levels along two distinct dimensions: (a) ability to learn from prediction errors, and (b) sensitivity to prediction errors.

Utilization and Dissemination of Findings:

The DOT task allows us to explore the differences between participants who excel at active inference and participants who do not. We plan to continue this research effort to hone the DOT task's efficacy at assessing Army officers' strengths and weaknesses in navigating and making sense of dynamic and uncertain situations. By learning about the factors that lead to efficient active inference, the Army is better positioned to develop methods to accelerate acquisition of competencies among Army officers to support decision-making under uncertainty.

ACTIVE INFERENCE: A COMPETENCY FOR MAKING DECISIONS IN UNCERTAIN SITUATIONS

CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACTIVE INFERENCE: A COMPETENCY FOR MAKING DECISIONS IN UNCERTAIN SITUATIONS

Army officers make decisions within operational environments composed of complex interdependencies that are evolving over time. Given this situation, officers must make decisions while relying on information about their environment that is often incomplete. To be effective decision-makers, officers develop and revise their situational understanding from moment to moment, inferring patterns from past and present events to anticipate probable future events (Stothart et al., 2023). Working with incomplete information, officers always perceive, decide, and act under some degree of uncertainty. Moreover, officers are themselves causal agents, contributing to the complexity and ambiguity of the environment in which they are acting. Every action they take has consequences that serve to further shape their future options. Their understanding of the operational environment is changed every time they decide, act, and then observe the consequences of their actions. When officers observe the consequences of their actions, it serves to shape their mental model of the environment. This revised mental model informs what they will do next.

What we describe here may be broadly conceptualized as a transactional process between a person and their environment, developing over time, and serving to shape perceptions, interpretations, decisions, and actions (see Lewin, 1943). To support this transactional process, and make informed decisions, officers must be capable of inferring meaningful patterns from ostensibly ambiguous information. That said, most training that officers receive focuses on the decision itself rather than the perceptual, inferential, and meaning making processes that encompass decision making over time (Graves et al., 2017). One reason for this is that the competencies that encompass decision making over time are very difficult to define operationally. Likewise, they are also difficult to measure. Without a clear operational definition and measurement model, producing and validating developmental interventions is nearly impossible. Consider, for example, a decision-making model often cited within the military decision-making literature: Colonel John Boyd's (1996) Observe, Orient, Decide, and Act (OODA) loop. The OODA loop is valued for how it describes an iterative decision-making process within situations characterized by volatility, uncertainty, complexity, and ambiguity (VUCA; Dempsey, 2012). While Army PME instructors acknowledge that the OODA loop presents a viable model for defcapacision-making, they have criticized it for challenges that arise when it is applied to train decision-making skills (Bartholomees Jr., 2010). The OODA loop model does not specify the competencies required to execute it nor is there a currently viable method to assess whether a decision-maker is performing the OODA loop effectively.

As an incremental step toward providing a solution to this problem, we focus on a psychological construct that may serve to encompass decision making over time—i.e., active inference—and provide a means to assess it. We present an assessment tool to diagnose individual officers' varying approaches to perceiving patterns in ambiguous information and a way to measure their performance. This tool may be valuable in training a critical aspect of how officers make operational decisions, as it provides a means for trainers to identify individual differences in approaches to active inference and to tailor training according to individual developmental needs.

When active inference first appeared in the cognitive neuroscience literature, it was described as a mechanism by which the mind seeks to resolve uncertainty and minimize surprise (i.e., if an outcome is highly surprising to an individual, then his or her beliefs might be generating invalid predictions and should therefore be revised; see Friston, 2005; 2009; 2010). When engaged with evolving operational problems, officers constantly compare their expectations for predicted outcomes to the actual outcomes of their decisions and actions. This observational process shapes ongoing decision making because operations rarely evolve exactly as predicted, requiring officers to revise and modify their situational understanding over time. With each iteration of prediction, decision, action, and outcome, the process begins again with revised beliefs that generate new predictions. In developing a measure of active inference, we intend to provide a means by which competencies encompassing decision making over time, and under uncertainty, may be assessed and enhanced.

Prior research indicates that Army officers differ in how they make decisions under uncertainty (Shortland et al., 2019). Some officers may exhibit great flexibility in managing a changing environment on a moment to moment basis, identifying potential patterns, anticipating future events, making decisions, and iteratively evaluating and adapting their decisions over time. Other officers may struggle with rigid thinking, persisting with a misaligned understanding of an evolving situation, trying to fit a preconceived—and potentially inaccurate—concept of the situation to their mission. When faced with persistent ambiguity and incomplete information, Army officers need to identify and select from many possible, yet equally reasonable, ways to respond to their evolving situation. While the OODA loop describes an approach by which flexible officers could find and apply different, but equally effective responses, there is no current method to assess the underlying competencies that may serve to encompass a decision (or a series of decisions) over time and within the iterative process the OODA loop model has proposed (Bartholomees Jr., 2010). We leveraged perspectives examining active inference to build just such an assessment: the Decisions Over Time (DOT) task.

Active Inference: The Context of the Research Problem

It is difficult to imagine how Army officers develop decision-making expertise for ambiguous and evolving environments. That said, situational contexts—even those that can be described as ambiguous and evolving—are an amalgam of invariant and varying features. Expertise typically develops as an individual acquires a variety of experiences in a domain, enabling them to derive knowledge about its invariant features (i.e., the patterns, meanings, or themes) that are in common across otherwise different experiences (Kahneman & Klein, 2009). When experts encounter situations relevant to their expertise, they can use their acquired knowledge to predict how situations will evolve much better than novices. Novices tend to mentally simulate multiple courses of action and intentionally compare them, whereas experts tend to recognize solutions almost immediately, as if guided by some unnatural intuition (Kahneman & Klein, 2009). Experts possess explicit and tacit knowledge about the patterns of meaning that give a conceptual form to the situations they encounter within their domain of expertise. Novices, who lack this explicit and tacit knowledge, are less equipped to predict how an unfamiliar situation will unfold. Therefore, novices are less able to solve the problem they are facing whereas an expert can readily perceive the relevant pattern that leads to a solution. However, expertise is a domain-contingent phenomenon, meaning there are occasions when an expert in one domain finds themselves in a strange and unfamiliar situation in which they are again a novice. The purpose of this project is to explore the underlying competencies that enable an individual to make sense of just such a strange and unfamiliar situation and to begin working toward developing the explicit and tacit knowledge required for expertise.

We can assume here that when people possess expertise, they are better able to adapt to situations that are relevant to their domain of expertise. Consider an example of this type of adaptation in the context of the Mann Gulch fire of 1949. A team of wildfire firefighters, led by Wagner Dodge, found themselves trapped by a wall of fire (Maclean, 1992). Dodge, realizing the life-threatening situation they were in, set fire to the tall grass around the team, and then laid in the ashes. By burning the fire's fuel before the fire reached him, Dodge created a small pocket of safety and the fire passed safely around him. Unfortunately, none of the other firefighters understood his plan and chose to run instead. Most did not survive. This idea, to create an escape fire, did not exist before Dodge came up with it. Ever since Mann Gulch, it has become part of all wildfire firefighting training. Although the invention of an escape fire is a prime example of expert-level adaptation, it was still a domain-specific adaptation. Dodge may have never faced the exact confluence of wildfire dangers he encountered that day, but he was an expert in his domain. Dodge's expertise allowed him to understand the flammability of tall grass, the difficulty of running uphill, the speed with which fire could travel, and many other small details that he could readily integrate to find a solution.

Since no one can be an expert in all domains, we each will encounter new situations that seem ambiguous and strange. How do we make effective decisions in those strange situations where we have become novices, again challenged to perceive the relevant patterns in the information we encounter? To understand how an expert in one domain navigates a strange situation in another domain, we must consider the process by which we extract meaningful information from any situation. Active inference describes the interplay between perception, cognition, and action necessary for acquiring an understanding of a new, strange situation (Friston, 2005; 2009; 2010). Friston, in his research on brain function, argued that the fundamental activity of the brain is to infer underlying causes for changes in sensory input. He proposed active inference as a process by which inferences that do not accurately predict sensory input are identified and revised, minimizing the likelihood of future surprise.

When engaged in active inference, individuals hold initial beliefs about an environment. These beliefs are used to predict potential consequences for specific actions (e.g., if I press the power button on my computer, the computer will turn on). After acting on a belief, individuals observe the actual outcome of their action, and they may recognize that the outcome they predicted is different from the outcome they observed (termed *prediction error* or *surprisal*). A large enough prediction error indicates to an individual that their beliefs are inaccurate or insufficient, since they experienced an outcome they did not anticipate. In less scientific terms, you learn by failing, figuring out why you failed, and then trying again after accounting for your previous mistake. Ideally, recognizing an error in how we understand the world should motivate us to revise our beliefs about the world, updating the belief we discovered to be incorrect. Through a transactional process, active inference leads to updated beliefs and those new beliefs serve to generate new—and potentially more accurate—predictions (Clark, 2013; Hutchinson & Barrett, 2019).

When comparing active inference to the OODA loop, several similarities become apparent. Both concepts describe a time-based process of perceiving, thinking, and acting that relies on information about outcomes to inform each subsequent iteration, with each movement through the process expected to improve the next outcome. There is more to this comparison, however. Behavioral scientists have long argued that individuals develop and adapt their predictions (Thorndike, 1927; Kahneman & Tversky, 1973; Rescorla & Wagner, 1972; Sutton & Barto, 1981; Holroyd & Coles, 2002). What makes active inference different is the way in which it frames the purpose of prediction. The traditional argument is that prediction is beneficial because it better prepares an individual to produce a desired outcome. But this traditional argument places the emphasis of prediction on actions. Comparatively, active inference argues that "a primary function of the brain is to infer the true state of the world in order to determine which motor behaviors will best promote adaptive fitness" (Fiorillo, 2010, p. 605). Instead of emphasizing the prediction of actions, active inference also emphasizes that inferred beliefs about the world are what elicit a prediction in the first place. In other words, you need to be able to perceive and think about the world accurately before you can act within it to produce results that will have value to you. Unlike the perspective that derives from reinforcement learning, active inference is not about learning an ideal response; active inference is about refining over time a mental model that will become increasingly accurate in guiding actions in an environment. It is a mechanism by which we develop principled knowledge about the world over time by being sensitive to and adapting our beliefs within situations in which our current ability to predict what will happen next comes up short.

Active inference describes a process by which we seek to minimize surprise. If an outcome is highly surprising, then our inferred mental model may be generating invalid predictions and should be revised. When someone has performed active inference extensively within a particular context, they may more accurately predict future outcomes because their mental model has already been extensively fine-tuned so that it better reflects the true state of the world. It seems reasonable then that active inference may also apply to ambiguous and novel situations where past experiences may be a poor guide to future behavior. This premise is because active inference enables individuals to probe whether their existing understanding of the world is serving as a valid guide to make sense of their current situation. The more efficient an Army officer is at active inference, the more prepared the officer will be to enter a situation they have never encountered before, to realize the limits of their current experience, and to begin the process of inferring a refined mental model to guide decision making in the novel environment.

How then do we best prepare an Army officer to make decisions in warfare when the operational environment is defined by volatility, uncertainty, complexity, and ambiguity? One potential answer is that the ideal officer will adapt to and perform effectively in almost any situation, familiar or unfamiliar. In an unfamiliar situation, however, prior experience can be detrimental to decision making, as prior training and experience may cease to provide an accurate mental model. Even so, it does not make sense to write off the extensive training and experience that officers accrue over a career. Perhaps a better answer to the question is that an ideal officer would be flexible in how they use their existing knowledge and experience, adapting what they know to make sense of what they do not yet know. An officer needs to be ready to revise their existing beliefs when facing a result they did not predict (Shadrick et al., 2007). In addition, they must be able to quickly generate a mental model that can serve to guide their decisions and actions.

The present research presents a way to assess an Army officer's ability to infer new knowledge about an ambiguous situation by tracking how they iteratively construct and update their present understanding based on the observed results of their choices and actions. Once we can define and measure what exemplary active inference looks like, we will be better positioned to enhance Army officers' capacity: (a) to identify the ways in which their experience may limit their understanding, and (b) to actively infer new knowledge within situations for which they have no existing knowledge and experience.

Developing an Assessment

Several challenges arise when determining how to assess active inference. For instance, some researchers have argued that fluid intelligence and other IQ assessments (e.g., Raven's Progressive Matrices, Wechsler Adult Intelligence Scale) assess the ability to perform abstract predictions (Euler, 2018), but these IQ assessments do not require participants to learn from experience or outcomes. The measures rely on a one-and-done set of sequential problem-solving tasks, the outcome being a correct or incorrect response to the question presented. Participants complete a series of complex abstract reasoning problems that are entirely independent of one another. Moreover, these fluid intelligence measures often provide all necessary information to figure out the answer, with the primary difficulty being the complexity of the mental computations required to do so. For decision making under uncertainty, such an assessment model is conceptually wrong. Life rarely presents us with independent problems, containing all the information we need to solve them. To address real world problems, we grapple with an environment full of hidden interdependencies—working with incomplete and imperfect information—trying to make sense of things as they evolve in real time. We deal with those situations as best we can because we rarely, if ever, find the perfect solution the first time we try. For Army officers, there is no one-and-done approach to operational decision making.

Consider for a moment two perspectives on decision making that are pertinent to active inference: probabilistic learning and executive functioning. Both perspectives highlight complex cognitive processes to varying degrees. While probabilistic learning focuses on how risk and uncertainty implicitly influence decision making (Damasio, 1996), executive functioning focuses on how problems are evaluated logically, how problem-solving approaches are planned and executed, and how hypotheses are tested and refined—with all these activities serving to identify the best solution to a problem (Chan et al., 2008). Another difference between these perspectives concerns the proposed role of emotion and cognition. For instance, probabilistic learning emphasizes emotion and arousal in decision making (Buelow & Suhr, 2009) whereas executive function emphasizes mechanistic cognitive operations, such as shifting mental sets, maintaining and updating working memory, and inhibiting prepotent responses (Miyake et al., 2000).

Probabilistic learning and active inference both emphasize Bayesian probabilities (Friston, 2008). Several published measures were designed explicitly to elicit Bayesian reasoning, i.e., the type of reasoning a person engages in when working incrementally toward a solution to a problem about which they have only incomplete information. Two examples are the

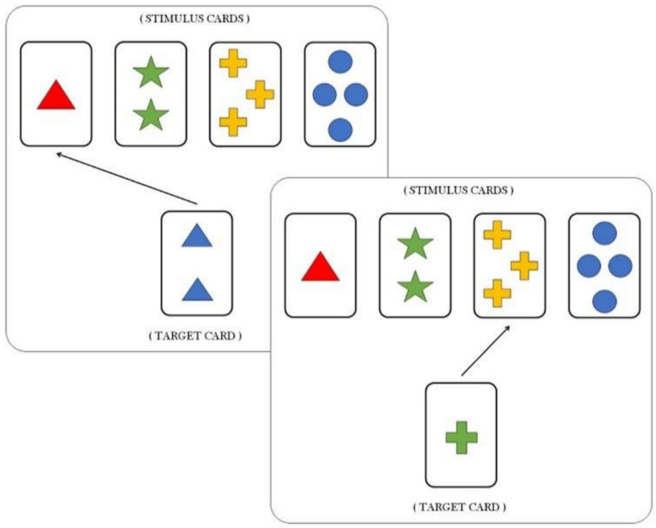
Multi-Armed Bandit Task and the better-known Iowa Gambling Task (IGT; see Damasio, 1996). In such tasks, participants are presented with four choices, often represented by decks of cards, that are denoted with neutral terms like an alphanumeric title (e.g., A, B, C, or D). After selecting a deck, participants are presented a number, generally intended to represent their monetary gain or loss. Each deck has a preestablished probability of producing a particular gain or loss. Optimal performance in such a task requires an individual to learn to estimate the hidden probabilities through feedback about their gains and losses alone, and to revise their choices as the probabilities change. Although significant debate continues concerning the core constructs being measured (Toplak et al., 2010), these tasks have found clinical application in assessing challenges related to managing risk and uncertainty (Buelow & Suhr, 2009). However, when looking to these measures as models for how we could approach assessing active inference, we identified a limiting factor: no higher-order rule was present to determine the probabilities of the decks. Participants simply had to learn implicit response-outcome associations with each deck and to respond according to those associations.

To better operationalize a Bayesian reasoning process, we needed participants to infer a latent rule that could be used to guide all hypothesis-testing behaviors within the task. One such task we identified was the *Wisconsin Card Sorting Task* (*WCST*; Grant and Berg, 1948; see Barceló, 2021 for an active inference interpretation). This task has extensive history in the executive function literature and is considered ideal to assess how well an individual can adaptively shift between mental sets (Miyake et al., 2000).

In the WCST, participants view five cards, with four cards across the top of the screen (category cards) and one card displayed in the center (target card; see Figure 1). Each of the category cards contain symbols that vary in number, shape, and color. Participants are told to place the target card on the correct category card but are not given any further information. After placing the target card on any category card, the participant is told whether they are right or wrong. To succeed, participants need to engage in active inference by developing hypotheses about the task, testing their hypotheses with an action, and then updating their beliefs based on the observed outcome (Barceló, 2021). Although these beliefs are uncertain at first, each observed outcome can be used to resolve further uncertainty, all depending on how effective the participant is at engaging in active inference across trials. Over time, the participant must learn that the four category cards represent four instances of three different concepts (i.e., number, shape, color), and that the sorting rule is related to only one concept at a given time. For example, if number is the sorting rule, then participants should sort a target card to the category that matches its number regardless of its shape or color. However, what participants do not know is that the rule changes after a certain number of consecutive correct matches, requiring participants to adapt. The WCST is a continual assessment of individuals' ability to resolve uncertainty by interpreting the outcomes of their actions, including when their knowledge becomes irrelevant, requiring them to inhibit outdated knowledge, and to adapt accordingly to a changing environment. Unlike the IGT, which requires participants to learn individual response-outcome associations, the WCST requires participants to infer a higher-order concept to generate and test subsequent hypotheses about the rules governing optimum performance.

Figure 1

Wisconsin Card Sorting Task Example



Note. Example trials from the WCST. In both trials, the participant is sorting according to the "shape" rule as opposed to the "number" rule or to the "color" rule.

Unfortunately, the WCST is not designed for normal healthy adult populations. It is primarily used to assess executive function deficits among children, older adults, or clinical populations (e.g., assessing for stroke, traumatic brain injury, dementia) and it is very rarely used with normal adults (Kopp et al., 2019). The likely reason for this rare use is that the WCST has a ceiling effect with normal adults, as the performance of healthy adults is relatively indistinguishable from that of normally developing adolescents (Chelune & Baer, 1986; Welsh et al., 1991). We propose several reasons for this ceiling effect. First, the potential rules are strongly cued, with each card consisting of highly distinct symbols. Second, four distinct choices are always presented, which likely assists participants in generating hypotheses and comparing outcomes across trials. Finally, participants must learn only a single rule at a time while normal adults can usually hold 3–5 items in working memory at a time (Cowan, 2010). The requirement of inferring a single rule likely does not elicit a high level of active inference ability. Therefore, the WCST is likely inappropriate as a measure to identify high performers (Diaz-Asper et al., 2004).

Decisions Over Time (DOT) Task

When both the IGT and WCST seemed ill-suited to our purpose, we identified a task that had been used by Friston and colleagues (2017a) to assess a computational model of active inference. They had used the task to model optimal learning based on feedback following an action, but they had not used the task to explore human performance. We incorporated the adaptive component from the WCST and then modified Friston and colleagues' task to assess active inference in human populations. This task, which we call the Decisions Over Time (DOT) task, operates according to the following parameters:

1) Success in the task is defined by a specific set of rules that are initially unknown to participants.
2) After making a choice, participants are told whether it was correct or was incorrect.
3) After ideal performance is achieved, the rules change, requiring participants to inhibit the outdated rules and to learn new rules.

The DOT task requires participants to learn a complex structure of rules that govern a process they have never encountered before. The only way participants can learn the rules is by making choices and inferring from outcomes, eventually developing beliefs about the structure of the multiple rules that enable ideal performance. The primary difference between the DOT task and Friston et al.'s (2017a) is that the DOT task requires adaptation. Friston and colleagues were primarily interested in testing how rapidly an active inference computational model could achieve ideal performance, and therefore did not require an adaptation component. We intended to extend Friston and colleagues' research by applying the DOT task to assess adaptation when rules become outdated, thereby creating a task analogous to the WCST but also capable of assessing high and low levels of active inference ability.

With a means to measure active inference, the Army will be better positioned to assess and develop a competency officers require to make effective decisions over time in an ambiguous environment about which they have incomplete information. Such a tool may further enable the Army to develop individually tailored training interventions to accelerate acquisition of Army officers' active inference skills.

Method

Participants

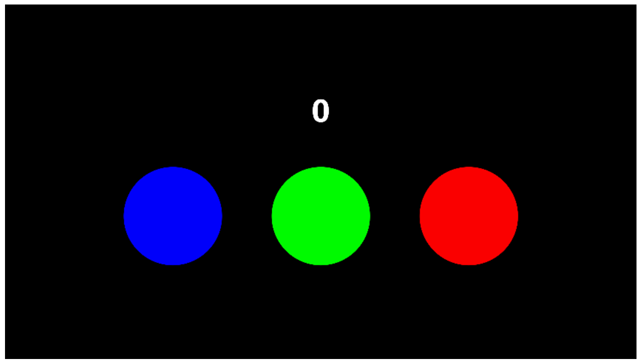
Participants were 51 noncommissioned officers (NCOs) and commissioned officers, drawn from Army installations across the United States.

DOT Task

In this computerized task, programed in JavaScript, participants were shown three dots (1 red dot [R], 1 green dot [G], and 1 blue dot [B]) arrayed in a row (see Figure 2). The sequence of colored dots varied across trials with the constraint that no two dots shared the same color during a trial. This approach resulted in a total of six distinct sequences of color combinations, also referred to as items (BGR; BRG; GRB; GBR; RGB; RBG). During each trial, one of the dots was made the correct dot and the other two dots were made the incorrect dots. Participants were instructed to click on the correct dot. The software chose which dot was correct according to a rule. Each rule was based on the color of the middle dot, which we referred to as the key dot. When a specific rule was active, the color of the key dot determined the location of the correct dot. As there were three different colors indicating three different correct locations, each individual rule was composed of three distinct response contingencies. For example, one rule consisted of the following response contingencies: If a green middle dot, then the rightmost is correct; if a red middle dot, then the leftmost is correct; if a blue middle dot, then the middle is correct (see Figure 3). To assess adaptability, we programmed the task so that the rule changed after participants successfully scored 5/6 trials correct for two consecutive blocks. Each block contained 6 trials and no sequence of dots was replicated within a block (see Figure 3). To do well, participants needed to form and to subsequently update their predictions about the rule.

Figure 2

DOT Task Example

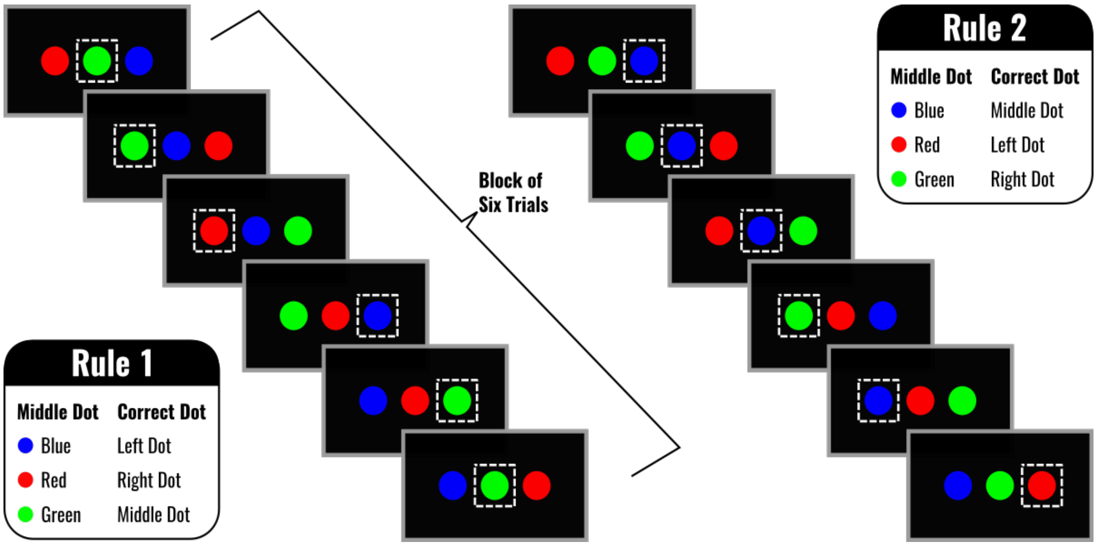


Note. Screenshot of the DOT task. The three dots represent potential choices and the white number represents the participant's current score.

To motivate participants, we gamified the task such that participants scored a point if they clicked on the correct dot and they lost a point if they clicked on an incorrect dot. Participants were shown their cumulative score above the middle dot. We applied this design as gamification has been shown to increase task motivation (Hamari et al., 2014). Additionally, participants were instructed that if they performed well, the task would end early. The task ended when participants achieved 40 points or after 20 minutes.

Figure 3

Example of DOT Task Trials



Note. Example of DOT Task Rules. Dashed boxes represent correct answers. Dashed lines do not appear in actual task.

Self-Report Measures

Background Survey

We used a background survey to collect basic demographics, such as rank, years of service, education, etc. We also assessed past/present hobbies and Army position history (e.g., platoon leader).

Ten Item Personality Inventory (TIPI)

The TIPI (Gosling et al., 2003) is a 10-item measure of the Big Five personality dimensions. Participants responded to each item (e.g., "I see myself as extraverted, enthusiastic") using a 7-item scale (1 = *disagree strongly* to 7 = *agree strongly*).

NASA Task Load Index (NASA TLX)

The NASA TLX (Hart & Staveland, 1988) is a 6-item scale that assesses respondents' perceived workload from doing a task. The 6 items on the scale assess mental demand, physical demand, temporal demand, effort required, perceived task performance, and frustration with the task. Participants responded to each item using a 20-item scale (1 = *very low* to 20 = *very high*).

Propensity to Perceive Coincidences Scale (PC)

The Propensity to Perceive Coincidences Scale (Bressan, 2002) assesses the frequency with which participants perceive coincidences. This scale assesses the degree to which participants possess "loose" cognitive control, for instance, less rigid associative networks and reduced inhibition of irrelevant memory content. The scale has 14 items. For each of the first 7 items, participants indicated how frequently they encountered certain coincidences (e.g., "perception of something distant in time, like having a dream that comes true") using a 5-item scale (1 = *never* to 5 = *very often*). This subscale is referred to as PC-frequency. Using the last 7 items, participants indicated their belief in the meaning and reasons behind coincidences (e.g., "do you think coincidences are due to pure chance?") using 3 response options (*Yes*, *No*, *I don't know*). This latter subscale is referred to as PC-meaning, with *yes* responses being scored as 1, *no* responses being scored as 0, and *I don't know* responses being scored as 0.5.

The Mindful Attention Awareness Scale (MAAS)

The MAAS (Brown & Ryan, 2003) is a 15-item scale that assesses a core characteristic of mindfulness: a receptive state of mind in which attention, informed by a sensitive awareness of what is occurring in the present, simply observes what is taking place. Participants responded to each item on the scale (e.g., "I could be experiencing some emotion and not be conscious of it until sometime later") using a 6-item scale (1 = *almost always* to 6 = *almost never*).

Procedure

Data collection was conducted in approximately 1-hour sessions with up to three participants in each session. Participants first completed the collection of surveys, then they performed the DOT task, and finally they responded to the TLX based on their experience performing the DOT task. We scored performance on the DOT task in terms of total score, total number of rules, accuracy, and average response time (RT) in milliseconds (ms). We capped the total score at 40 because that was the cutoff for task completion, but it could become negative if participants responded incorrectly more than they responded correctly. There was no limit to how negative the score could become. We calculated the total number of rules as the number of rules to which participants responded.

Results

Of the 51 participants who consented to participate, we excluded the data of four participants for the following reasons: one participant opted out shortly after consenting to participate; data from two participants were lost due to computer error; and we identified one participant as low effort on the DOT task as this participant selected the right-most dot on more than 99% of the trials regardless of receiving negative feedback through a diminishing score. Therefore, we analyzed the data of a total of 47 participants. Descriptive statistics on demographic variables for this sample are presented in Table 1. The aggregated survey data consisted of 366 data points from the 47 participants, with 10 missing data points, totaling 2.6% of the data. These cases were ignored when those data were required for analysis, as there was insufficient relevant data for imputation. All data were analyzed using the R statistics program (R Core Team, 2013).

Table 1

Sample Demographics

Variable	NCOs ($n = 21$) Mean	Officers ($n = 26$) Mean
Age	37.62	31.52
Years in service	15.90	8.65

Intercorrelation coefficients across the survey measures are presented in Table 2. Missing data were handled through pairwise-deletion of cases. Results show moderate correlations among the personality variables Openness, Agreeableness, and Emotional Stability ($r = .33–.37$) as well as a moderate correlation between MAAS and Emotional Stability ($r = .43$). Finally, there was a moderate correlation between the PC subscales frequency and meaning ($r = .34$).

Table 2

Correlation Matrix of Survey Measures.

Variable	MAAS	PC Freq.	PC Mean.	Extra.	O	A	C	E
MAAS	—							
PC Freq.	-.04	—						
PC Mean.	.01	**.34**	—					
Extra.	.18	.06	-.12	—				
O	.14	.15	.18	.23	—			
A	.04	.00	.12	-.01	**.33**	—		
C	.27	.08	.04	.25	.04	.01	—	
E	**.43**	.11	-.05	.08	**.33**	**.37**	.20	—

Note. MAAS = Mindful Attention Awareness Scale, PC Freq. = Propensity to perceive meaningful coincidences: frequency subscale, PC Mean. = propensity to perceive meaningful coincidences: meaningfulness subscale, Extra. = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, E = Emotional Stability.
Bolded values are significant at the $p < .05$ level.

Although the DOT task was designed to assess individuals' ability to learn the specific rules, debriefing interviews with participants revealed that the DOT task could be executed using different strategies, which we referred to as *adapting* and *satisficing*. Adapters learned the specific rules and adapted to new rules when the initial rules changed. Satisficers found a statistical regularity in the task. For each rule, there was one color that was correct 66% of the time, specifically, the color that indicated the middle dot was correct (i.e., itself). As shown in Figure 3, that color would be correct both times that it was in the middle as well as once on each side, making it correct on 4/6 of the items. Since 66% accuracy would increase scores while also never eliciting a rule change, it was possible for satisficing participants to increase their score by simply responding to that color 100% of the time regardless of which item was being presented. They maintained this strategy given it was effective in increasing their score, while also controlling the point-loss penalty due to error.

After reviewing the data, we identified four distinct types of performers: adapters, overlearners, satisficers, and guessers. Adapters were participants who successfully elicited a rule change at least twice, meaning they responded to at least three rules. Overlearners were those who successfully learned the initial rule but failed to adapt and fully learn the second rule. Satisficers were participants who never fulfilled a rule change and adopted a strategy of selecting the same color dot for at least 5/6 of all items in a block. Finally, guessers were all other participants. The distribution of participants across the four performance groups was roughly equal (range: 10–13 participants in each group; see Table 3).

Table 3

Descriptive Statistics and Task Performance Across Groups

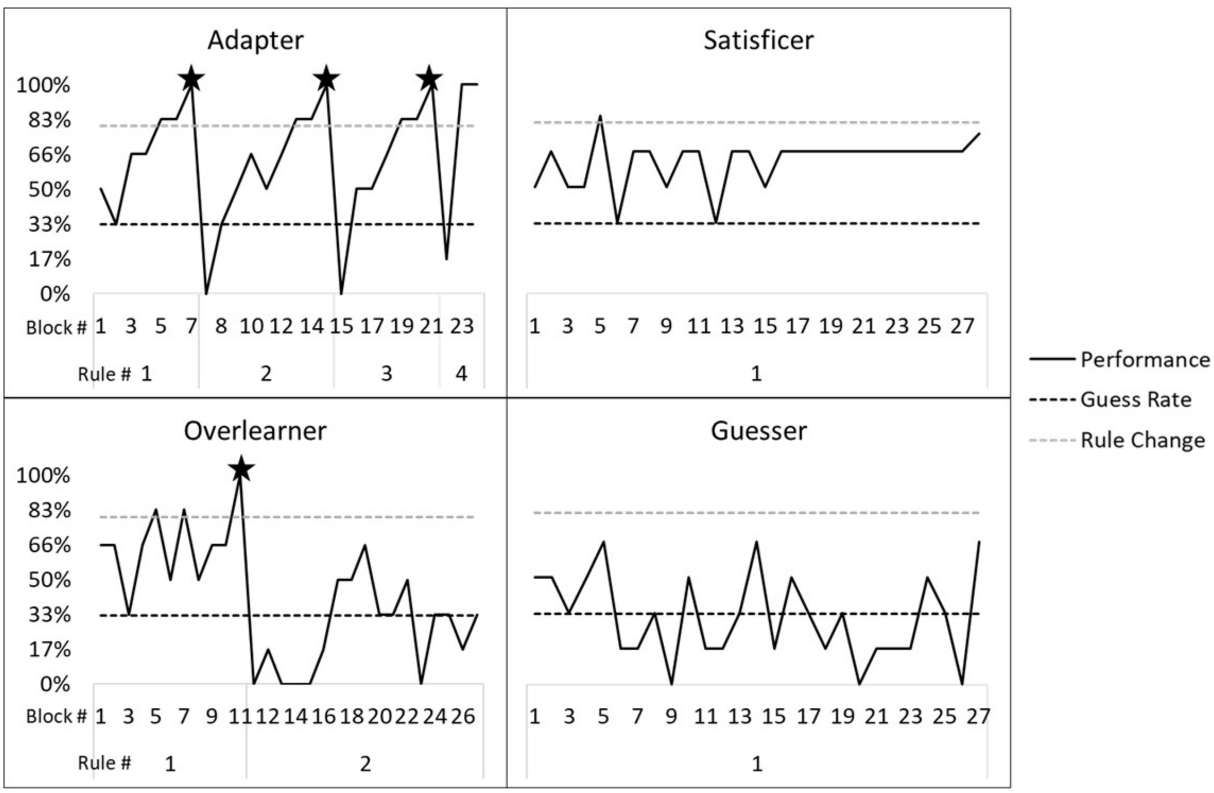
Group	Description	*n*	Mean Score (SD)	No. of Rules
Guessers	Failed to identify and act according to the rule.	12	-45.34 (33.80)	1 (0)
Overlearners	Identified a rule but failed to adapt to a new rule.	12	-6.92 (20.07)	2 (0)
Adapters	Identified the rules and adapted after the rules changed.	13	18.62 (31.87)	5 (1.64)
Satisficers	Identified an imperfect but good-enough probabilistic strategy.	10	37.30 (8.54)	1 (0)

Note. Shown are the means and standard deviations (in parentheses) of the final task score and of the number of rules learned.

An exemplar participant's data from each of these groups is presented in Figure 4. The data are accuracy rates (i.e., correct or incorrect) binned into each block of six items (see Figure 3 for more clarification on blocks). The *x*-axis has two distinct pieces of information. The top sequence of numbers refers to the block of trials in the order that the participant responded to them while the bottom number refers to the ordinal rule that the participant was currently responding to (e.g., the bottom number 2 means the participant had successfully solved the initial rule and the participant was now working on a second rule). The *y*-axis refers to that participant's accuracy rate in each block. For example, if the graph shows a 66% accuracy, it means that the participant responded accurately to 4 out of 6 items in that block. The black dashed line indicates the level of performance expected by chance alone, and the green dashed line refers to the accuracy threshold required for the task to change the rule. The green circles indicate the block in which this accuracy threshold was met.

Figure 4

Data from Single Exemplar Participant for each Group



Note. Data are accuracy rates binned into six-trial blocks. The y-axis depicts the accuracy rate of each block. For the x-axis, the topmost value is the ordinal number of each block while the lower number is the ordinal number of the active rule. The black dashed line depicts expected performance for pure guessing while the grey dashed line depicts the performance threshold to trigger a rule change. The task triggered a rule change if a participant's average performance over 12 trials met or exceeded the threshold. The Star symbol indicates a rule change.

A frequency table for task success is presented in Table 4, showing how effective the different groups were in completing the task, defined as achieving 40 points within the 20-minute time window. As shown in Table 4, no guessers or overlearners completed the task, whereas over half of the adapters and nearly all the satisficers completed the task. As shown in

Figure 5, the minimum amount of time required to complete to task appeared to be roughly 10-11 minutes for both satisficers and adapters. However, satisficers appeared more likely to complete the task than adapters, as shown in Table 4. We conducted a chi-squared test comparing adapters and satisficers across likelihood for time to expire or to complete the task. Although the effect of Group on task outcome was not statistically significant, $\chi^2(1, 23) = 3.49$, $p = .062$, $\varphi = 0.39$, there was a medium to large effect size, which is of sufficient size to merit reporting (Cohen, 1992).

Table 4

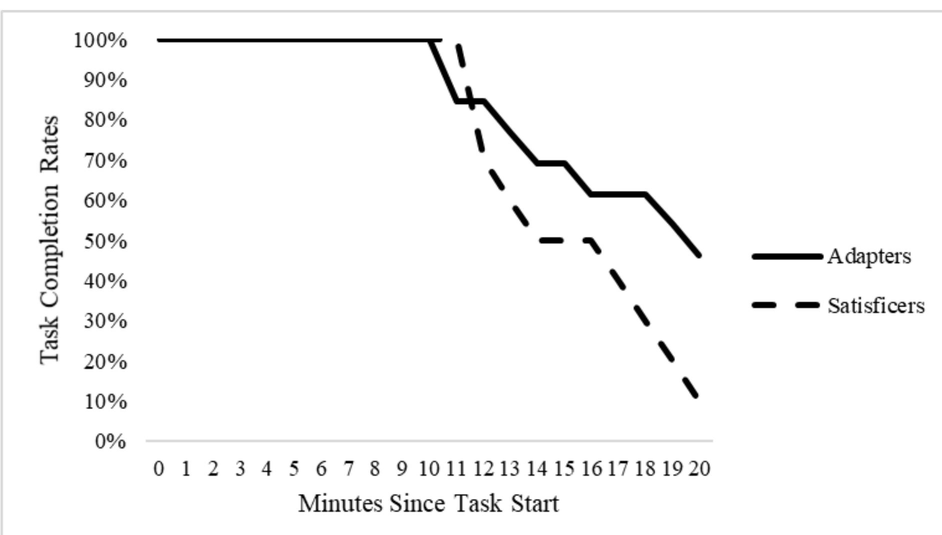
Task Completion Across Groups

Group	Time Expired	Completed Task	Total *N*
Guessers	12	0	12
Overlearners	12	0	12
Adapters	6	7	13
Satisficers	1	9	10

Note. Data are frequencies of participants where either their time expired or they completed the task. Task completion which is defined by the achieving of 40 points within 20 minutes of beginning the task whereas time expired means 20 minutes passed before achieving 40 points. Total *N* for each group is presented in the final column.

Figure 5

Completion Rate by Minutes Since Task Start



Note. Data are the rates of task completion, which is defined by the achieving of 40 points within 20 minutes of beginning the task. Only satisficers and adapters are presented here as no participants in the other groups completed the task. The y-axis states the completion rate while the x-axis presents the number of minutes that have passed since the task start.

Descriptive statistics for each of the survey measures can be seen in Table 5, broken out by group. We conducted a series of one-way Multivariate Analysis of Variance (MANOVA) and Analysis of Variance (ANOVA) tests to examine if the groups differed between the survey measures, using case-wise deletion to handle missing data. When a measure was part of a broader group of measures (e.g., five-factor personality), we conducted a MANOVA to test if there was a global effect of group that merited examining the origin of the effect.

Table 5

Survey Measures Across Group

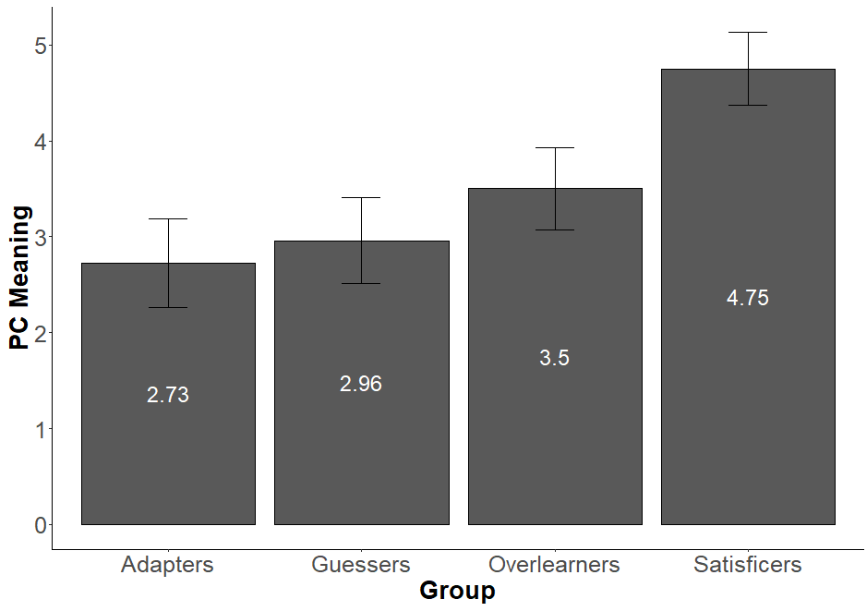
Variable	Guessers	Overlearners	Adapters	Satisficers
MAAS	3.97 (0.69)	4.62 (0.64)	4.12 (1.00)	4.12 (0.84)
PC Freq.	2.00 (0.68)	2.05 (0.70)	2.23 (0.59)	2.59 (0.68)
PC Mean.	2.96 (1.55)	3.50 (1.50)	2.73 (1.54)	4.75 (1.21)
Extra	3.69 (1.04)	3.96 (1.64)	3.70 (1.80)	4.15 (1.50)
O	5.37 (0.78)	4.71 (1.08)	5.00 (1.20)	5.25 (1.17)
A	4.78 (1.20)	4.21 (0.79)	4.31 (1.00)	5.45 (1.58)
C	5.46 (1.34)	6.09 (1.15)	5.93 (0.61)	6.00 (1.03)
E	5.60 (1.21)	5.46 (1.12)	5.16 (1.48)	5.65 (1.83)

Note. Shown are means and standard deviations (in parentheses). MAAS = Mindful Attention Awareness Scale, PC Freq. = Propensity to perceive meaningful coincidences: frequency subscale, PC Mean. = propensity to perceive meaningful coincidences: meaningfulness subscale, Extra. = Extraversion, O = Openness, A = Agreeableness, C = Conscientiousness, E = Emotional Stability.

We conducted a MANOVA on the personality variables, which was insignificant, $F(3, 41) = 0.81$, $p > .05$. We also found no effect related to the MAAS measure, as revealed by an insignificant ANOVA, $F(3, 42) = 1.43$, $p > .05$. We did find significant effects after conducting a MANOVA on the PC measures, $F(3, 41) = 2.29$, $p = .043$, $\eta^2 = .14$. Follow-up ANOVAs revealed no effect on PC-frequency, $F(3, 41) = 1.74$, $p > .05$, and a very large effect on PC-meaning, $F(3, 41) = 3.99$, $p = .014$, $\eta^2 = .23$. We probed the source of this effect by conducting post-hoc comparisons, controlling familywise error via the Tukey HSD correction. We found two significant differences, both related to the satisficers ($M = 4.75$, $SD =1.21$), who reported higher rates of PC-meaning than the adapters ($M = 2.73$, $SD = 1.54$, $p = .032$) and the guessers ($M = 2.96$, $SD =1.54$, $p = .015$). These data can also be seen in Figure 6. Finally, we conducted a MANOVA on the TLX variables, but found no effect related to group, $F(3, 41) = 0.96$, $p > .05$.

Figure 6

Bar Chart Depicting PC-Meaning Across all Groups



Note. The y-axis depicts the response on the Meaning subscale of the Propensity to Perceive Coincidences scale. The x-axis depicts the Group of participants based on their performance. Means are presented in center of bar graph. Error bars represent standard errors.

To evaluate how the different groups approached and performed the task, we examined group differences in Reaction Time (RT) as well as accuracy rates over the course of the entire task. To account for individual-level variability in problem-solving approach, Reaction Time and Accuracy were modelled using mixed-effects regressions (LMERs)[1]. We modelled all LMERs using the R packages lme4 (Bates, 2010) and lmerTest (Kuznetsova et al., 2017). Finally, we removed all trials where RTs were faster than 300 milliseconds, as these were unlikely to be relevant under choice RT conditions (Deary et al., 2011).

To examine differences in how the behavioral outcomes (RT and Accuracy) were related to group and to minutes since task start, we conducted a mixed-effects regression for each outcome, with variability in outcome as a function of group, coded as a between-subjects factor,

[1] LMERs allow for the modelling of both fixed and random effects as well as random slopes. Standard linear regressions assume that all observations are equally independent, which is not appropriate for a repeated-measures design. LMERs can account for dependencies within the observations that are likely to be correlated, but not in a manner that is easily predictable or relevant to the primary hypotheses. Such dependencies are incorporated as random effects. Similarly, LMERs can model random slopes within random effects, accounting for variance in predictors that is related to a random effect, further isolating variance that is uniquely related to the relevant predictor. The process of LMERs accounting for random effects is not entirely unlike how a within-subjects ANOVA models participants as a random effect, except LMERs can model multiple random effects simultaneously as well as random slopes within the random effects. For more thorough discussion of LMERs, see Baayen et al. (2008). For a theoretical discussion of the statistical concerns that LMERs are designed to address, see Clark (1973).

and with the number of minutes spent in task coded as a within-subjects factor. We mean-centered minutes by each participant. We constructed contrast codes for the fixed effect of Group, as can be seen in Table 6. The first contrast tested for differences between the groups that found successful strategies (i.e., adapters and satisficers) and the groups that did not find successful strategies (i.e., overlearners and guessers). The second contrast tested for differences between the successful strategy groups (i.e., adapters vs satisficers). The final contrast tested for differences between the unsuccessful strategy groups (i.e., overlearners and guessers). To account for idiosyncratic performance that was unique to individual participants and was unrelated to the constructs at hand, each participant was coded as a random effect and minutes elapsed from the start of the task was modelled as a random slope for each participant.

Table 6

Contrast Codes for Group

Group	Contrast 1	Contrast 2	Contrast 3
Guessers	-0.25	0	0.5
Overlearners	-0.25	0	-0.5
Adapters	0.25	0.5	0
Satisficers	0.25	-0.5	0

The first regression model examined RT as a function of group, minutes, and their interaction, as well as the random effects and slopes described above. Because the distribution of RT is often highly skewed and non-normal, it was log-transformed to better fit the assumptions of regression. *P*-values were estimated using Satterthwaite approximations (Luke, 2017). The regression results can be seen in Table 7 and a graph of the raw data, with a LOESS smoother applied, can be seen in Figure 7. The results revealed that the groups with a successful strategy responded more quickly than those with an unsuccessful strategy, and that within those groups with a successful strategy, satisficers responded more quickly than adapters. Although all groups responded more quickly as they proceeded through the experiment, as evidenced by the effect of minutes, the groups with a successful strategy showed a larger reduction in RT than the groups with an unsuccessful strategy. No other differences between groups were observed.

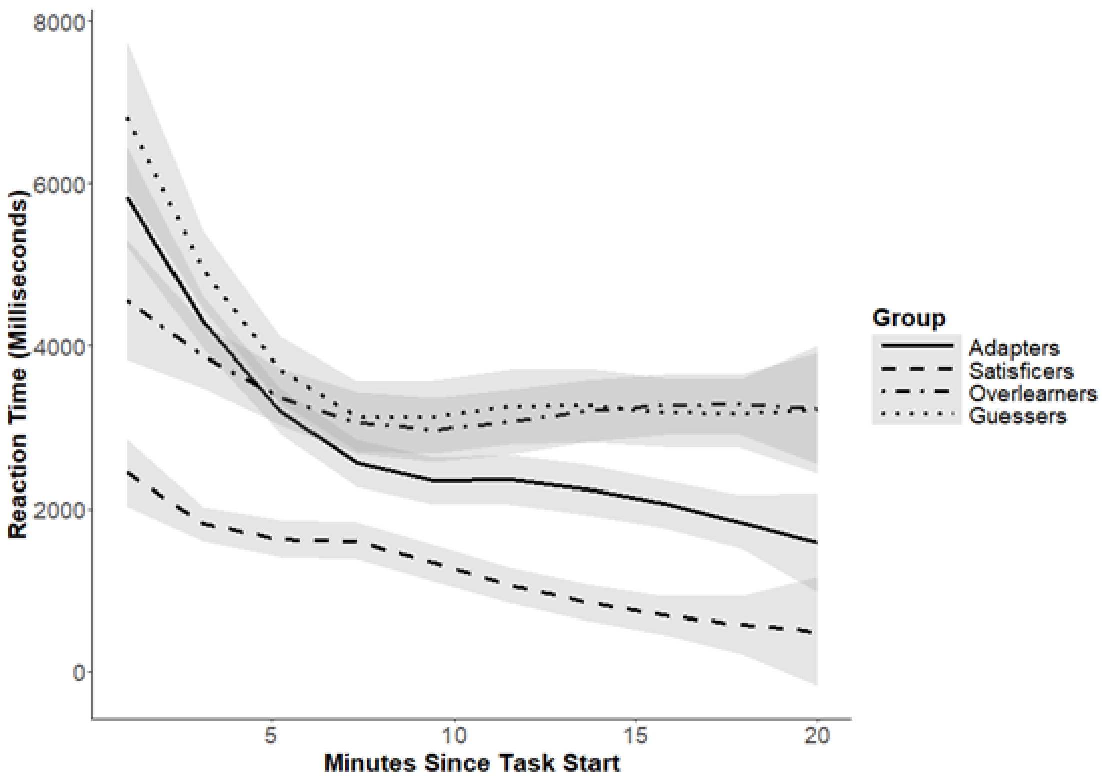
Table 7

Linear Mixed- Effects Regression Examining Reaction Time Across Group Over the Course of the DOT Task

Coefficient	Estimate	*SE*	*df*	*t*	*p*
(Intercept)	7.39	0.08	42.64	91.00	0.0000
Contrast 1: (Adapters + Satisficers) vs (Overlearners + Guessers)	-1.36	0.32	42.64	-4.18	0.0001
Contrast 2: Adapters vs. Satisficers	0.77	0.23	42.86	3.31	0.0019
Contrast 3: Overlearners vs. Guessers	0.00	0.23	42.41	0.02	0.9834
Minutes from start	-0.04	0.01	40.07	-6.07	0.0000
Contrast 1 x Minutes	-0.10	0.03	40.07	-3.61	0.0009
Contrast 2 x Minutes	0.00	0.02	41.78	-0.06	0.9499
Contrast 3 x Minutes	0.01	0.02	38.29	0.67	0.5091

Figure 7

Graph of Reaction Time Over the Course of the DOT Task

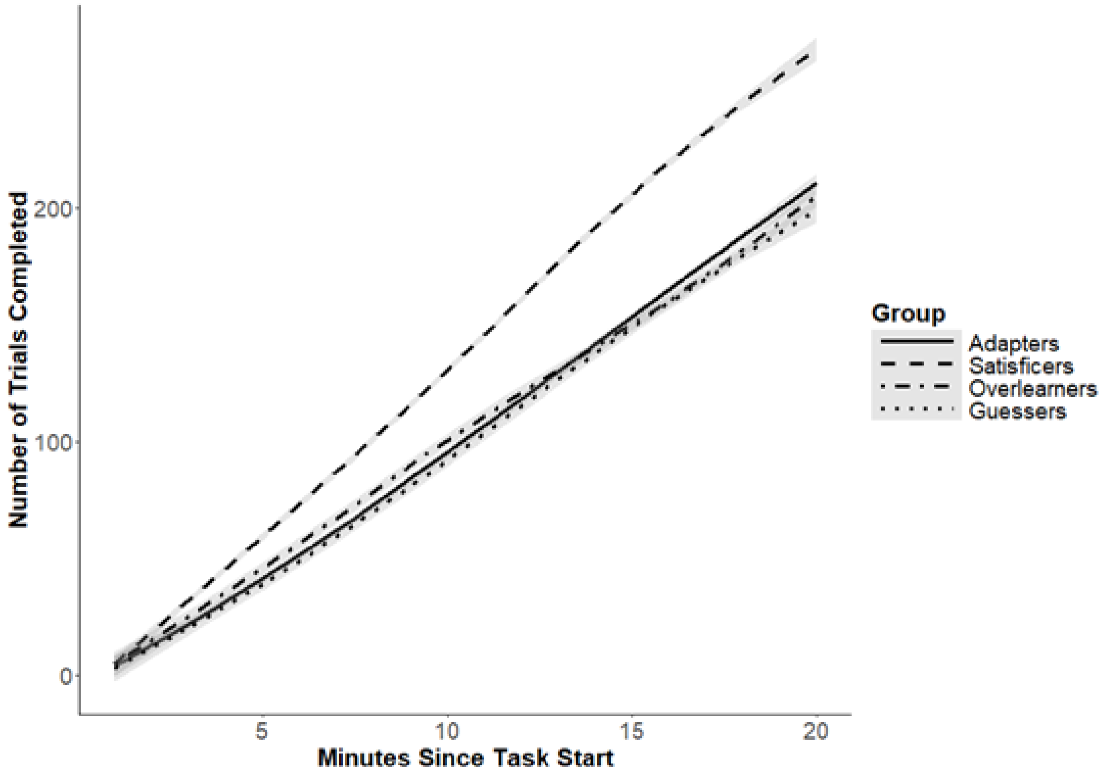


Note. The y-axis depicts the average reaction time for task response while the x-axis depicts the number of minutes since the task was started. Shaded region indicates 95% confidence interval.

Due to satisficers displaying faster reaction time than all other groups, and the task being self-paced, it seems reasonable to expect that satisficers experienced a greater number of trials than all other groups when the effect of task duration is controlled. Figure 8 depicts the average number of trials completed for each group by the number of minutes elapsed since participants started the task. This figure reflects the faster reaction time elicited from satisficers than all other groups, which was tested and verified in a standard linear regression analysis depicted in Table 8. While the average number of trials completed increased as task minutes increased for all groups, the satisficers exhibited notably higher rate of accumulating trials per minute than all other groups. The trial accumulation rate for adapters, guessers, and overlearners appeared to be roughly equivalent.

Figure 8

Number of Trials Experienced Across the Course of the Task



Note. The y-axis depicts the average number of trials completed while the x-axis depicts the number of minutes since the task was started. Shaded region indicates 95% confidence interval.

Table 8

Linear Regression Examining Number of Trials Experienced Across the Course of the DOT Task

Coefficient	Estimate	*SE*	*t*	*p*
(Intercept)	12.64	1.28	9.87	0.0000
Minutes from start	11.52	0.22	52.38	0.0000
Contrast 1: (Satisficers+ Adapters) vs (Overlearners + Guessers)	50.78	5.13	9.90	0.0000

Table 8 (*continued)*

Coefficient	Estimate	*SE*	*t*	*p*
Contrast 2: Satisficers vs. Adapters	41.62	4.01	10.38	0.0000
Contrast 3: Overlearners vs. Guessers	1.27	3.19	0.40	0.6920
Minutes x Contrast 1	4.91	0.88	5.58	0.0000
Minutes x Contrast 2	3.29	0.70	4.67	0.0000
Minutes x Contrast 3	-0.38	0.53	-0.72	0.4710

One concern with the above regression analysis is that there is a varying *N* size across minutes spent in the task due to a varying rate of task completion across the groups; that is, data on fewer participants is available as over time as individual participants complete the DOT task. For example, there are fewer satisficers in the task at minute 17 than there are at minute 9. This discrepancy complicates comparisons of trial count by minutes from task start and group. To account for this, we tested whether the greater number of trials completed by satisficers is maintained even when accounting for the effect of task completion. We isolated only the satisficers and adapters, and conducted an ANOVA, comparing the number of trials across Group (satisficers vs. adapters) and Outcome (Time Expired vs. Task Completion). This test largely confirmed the trial effect, as we found an effect of Group $F(1, 19) = 4.81$, $p < .05$ as well as an effect of Outcome $F(1, 19) = 4.77$, $p < .05$, with no interaction effect. These effects revealed a greater number of trials experienced by participants whose time expired (M = 230, SD = 61) than those who completed the task (M = 191, SD = 57) as well as greater number of trials experienced by satisficers (M = 226, SD = 52) than adapters (M = 185, SD = 60).

The third regression model examined accuracy as a function of group, time from task start (in minutes), the interaction of these variables, as well as the random effects and slopes described above. Given that accuracy is a binomial variable, we conducted a logistic regression to examine the probability of a response being *accurate* or *inaccurate* as a function of the predictors noted above. *P*-values were estimated using the asymptotic Wald test (Bolker et al., 2009). The regression results can be seen in Table 9 and a graph of the raw data, with a LOESS smoother applied, can be seen in Figure 9. The results revealed that the groups with a successful strategy responded more accurately than those with an unsuccessful strategy, and that within those groups with an unsuccessful strategy, overlearners responded more accurately than guessers. Although all groups responded more accurately as they proceeded through the experiment. As evidenced by the effect of minutes, the groups with a successful strategy showed a larger increase in accuracy than the groups with an unsuccessful strategy. Finally, among those groups with a successful strategy, adapters showed a greater increase in accuracy than satisficers, although the null effect of contrast 2 suggests that their overall scores for accuracy were relatively similar. No other differences between groups were observed.

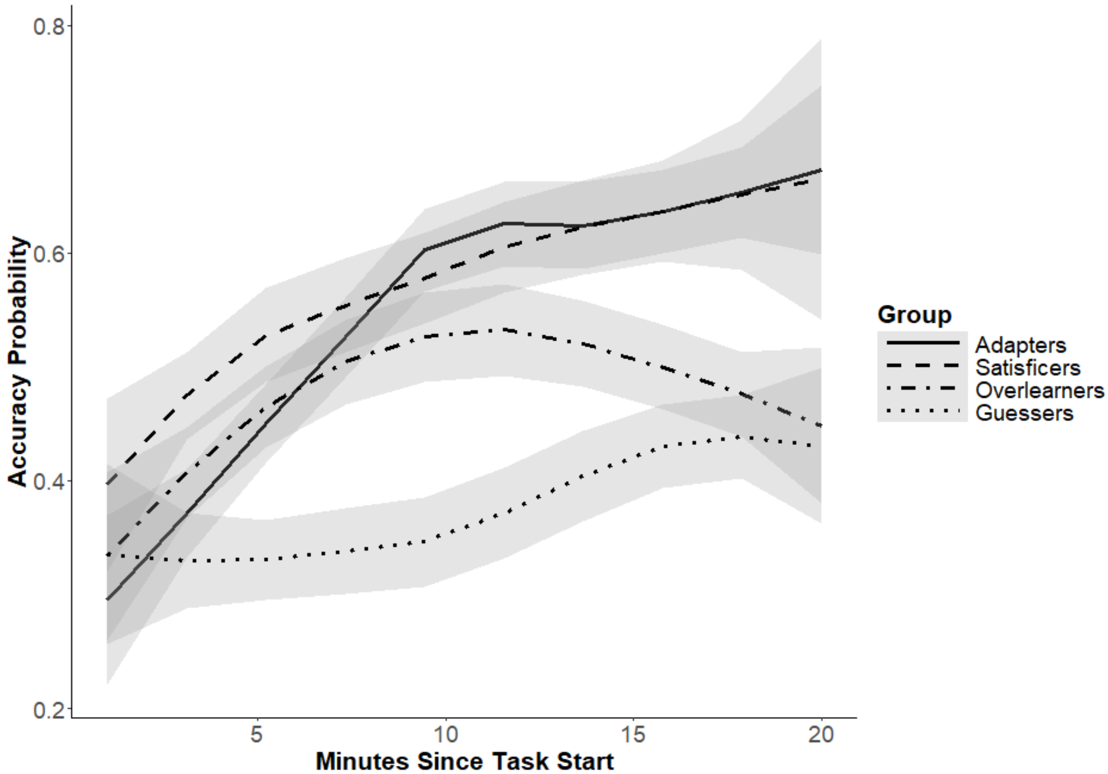
Table 9

Logistic Mixed- Effects Regression Examining Accuracy Across Group Over the Course of the DOT Task

Coefficient	Estimate	*SE*	*z*	*p*
(Intercept)	0.01	0.06	0.12	0.9046
Contrast 1: (Adapters + Satisficers) vs (Overlearners + Guessers)	1.32	0.22	5.89	0.0000
Contrast 2: Adapters vs. Satisficers	0.03	0.16	0.21	0.8347
Contrast 3: Overlearners vs. Guessers	0.48	0.15	3.16	0.0016
Minutes	0.06	0.01	6.42	0.0000
Contrast 1 x Minutes	0.13	0.03	3.85	0.0001
Contrast 2 x Minutes	0.06	0.03	2.30	0.0216
Contrast 3 x Minutes	-0.02	0.02	-0.85	0.3931

Figure 9

Graph of Accuracy Over the Course of the DOT Task



Note. The y-axis depicts the average probability of accurate responding while the x-axis depicts the number of minutes since the task was started. Shaded areas represent 95% confidence interval bands.

Discussion

The ability to infer latent patterns within situations that initially appear complex and ambiguous is a vital competency for effective decision making under uncertainty. The active inference construct presents a plausible account of the iterative perceptual process through which individuals' mental representations are tested and refined over time through predictions and interactions with an environment (Friston, 2009) and aligns well to existing decision-making models used in military contexts, namely the OODA loop (Boyd, 1996). The DOT task measures individuals' capacity to infer and revise their latent mental models through making decisions, acting, and interpreting the consequences of their actions (Friston et al., 2017a; Barceló, 2021). A particularly notable result of this research concerned the patterns of behavioral exploration that we identified. We initially intended the DOT task to be a high-level WCST, wherein the only way for participants to succeed was to learn the higher-order latent rules of the task. Even so, we discovered that we had inadvertently created a task that could be solved either through an executive functioning process (i.e., adapters) or a probabilistic learning process (i.e., satisficers). We interpret this serendipitous design flaw as a benefit, as it may enable this single task to assess individual preferences for two distinct—and successful—approaches to decision making under uncertainty: adapting and satisficing.

Before we consider results from the satisficers and adapters, it is important to note that the task demands for these groups were so different that it seems fair to suggest that these two groups self-selected into two different tasks. We initially developed the DOT task to assess the ability to learn a rule, apply it, then discard and revise it when faced with conflicting evidence. The satisficers never needed to revise their rules, which were much simpler than any of the rules learned and applied by the adapters. During debriefings, the lead author consistently observed that satisficers appeared energetic and excited to discuss the task whereas the adapters appeared exhausted and struggled to discuss the task. Although the omnibus ANOVA on the TLX was not significant, we did conduct specific exploratory *t*-tests within the TLX and found that adapters reported significantly greater Effort and Mental Demand than satisficers (ps < 0.05). Future research will need to correct for this disparity to ensure satisficing is still available as a viable strategy, but that those who pursue satisficing are pressured to revise and adapt their rules as well.

We designed the DOT task to assess active inference ability. Based on our interpretation of this literature and these results, we propose that the DOT task may assess active inference ability along two distinct dimensions: (a) ability to learn from prediction errors, and (b) sensitivity to prediction error. We believe there is sufficient face validity that the DOT task assesses some form of learning from prediction error, as the only information gained from the task was the interpretation of error-based feedback. For this first dimension, we argue that the guessers were the least effective at learning via prediction error, followed by the overlearners, and then both the adapters and the satisficers. We refrain from comparing the abilities of satisficers and adapters on this dimension because they were both relatively successful but were subject to different self-selected task demands (i.e., the satisficers applied a strategy that enabled them to bypass the need to adapt). Although one can make an argument that adapters learned more complex rules, the goal of the task presented to participants was to "click the correct dot," not "learn the rules of the task." Adapters and satisficers both accomplished this goal with similar overall accuracy.

Although we cannot yet differentiate between adapters and satisfiers in terms of efficacy at learning from prediction error, we hypothesize that satisficers possess more flexible thresholds of prediction error than adapters. According to active inference, an individual seeks to infer the causes of prediction error to enable them to better select actions that achieve goals (Linson et al., 2018). This means that individuals may possess a certain threshold of prediction error whereby individuals begin to revise their inferred models when that threshold of error tolerance is reached. Given this interpretation, we suggest that satisficers may possess more flexible error thresholds (i.e., higher tolerance for error) because they maintained their strategy while experiencing 33% error while adapters sought to pursue perfect accuracy throughout the task.

A second reason to interpret differences in error thresholds is that satisficers reported higher scores on the Propensity to Perceive Coincidences Scale, which is associated with a greater tendency to rely on internal predictive models and greater difficulty updating beliefs (van Elk et al., 2016; Corlett & Fletcher, 2012; Rominger et al., 2019). Previous applications of the

Coincidences Scale have tended to focus on tasks in which high scores on the scale are associated with impaired performance (which has been proposed to be related to reduced working memory and inhibition, see Rominger et al., 2011; 2019). We have been unable to identify research in which the Coincidences Scale was positively associated with higher levels of task performance. To make sense of this result, consider the core difference between the DOT task and most other cognitive tasks. Most cognitive tasks are precise in defining task success, presenting a well-defined problem and solution path. The DOT task was designed to focus on the process by which participants are working toward one solution among many, utilizing an ill-defined problem structure and solution path. Considering these differences in the character of task design, we suggest that the DOT task may be partially assessing individual tendencies to consider a broader range of possibilities for solutions and a greater acceptance of noise and uncertainty in the structure of the problem presented. The characteristics of the DOT task may have provided satisficers with greater flexibility to consider broader patterns across many trials instead of becoming driven by each individual outcome. Future research will be needed to validate the distinction we have made between adapters and satisficers as trait-level differences in tolerance for error.

This research also presents an opportunity to reconsider what an *optimal* strategy is. Studies that examine decision making under uncertainty generally have an ideal strategy already presupposed by the experimenter. This strategy is usually targeting an outcome focused on task performance, such as achieving a high score. However, this research suggests a different way to consider optimal strategy, that is, accomplishing the task while minimizing effort. Friston proposed active inference as a mechanistic process by which the brain seeks to make optimum inferences while minimizing energy expenditure (Friston, 2005). In our experiment, the satisficers produced a strategy that required taking more actions, (i.e., more trials), but appeared simpler to implement and less mentally taxing. Such a strategy appears like what Friston and colleagues describe as balancing pragmatic and epistemic value (Friston et al., 2017b). One could argue that the adapters eventually achieved more perfect performance than the satisficers at the individual trial-by-trial level. Even so, it is an open question whether the improved performance of adapters was worth the extra energy they expended to seek perfect solutions. In a context where loss of life could be the consequence of a bad decision, a decision-maker needs to get it right. Certainly, there are some situations where perfect performance is necessary, but there are many other problems where several good-enough solutions are satisfactory in moving toward a goal, with additional effort no longer providing tangible value (Manski, 2017). The DOT task paradigm may enable researchers to study the individual-level decision to prioritize either performance or efficiency in a manner that is subtle and not overt, enabling future research on factors that are fundamental to the management of uncertainty and to strategy switching.

Limitations and Future Directions

The first limitation is that we cannot currently interpret differences between adapters and satisficers. The active inference literature argues that some aspects of sensitivity to prediction error can be stable across a lifetime (Paulus et al., 2019) whereas other aspects can be quite malleable (Deane et al., 2020). Unfortunately, the current task is ill-equipped to examine differences between adapters and satisficers because satisficers never needed to adapt. They found a solution that worked and ran with it. It is theoretically possible that some satisficers simply started guessing and stumbled on this approach, making their success more of a fortunate accident than a skillful strategy. However, it is also possible that satisficers strategically extracted information from the environment and identified an effective strategy to achieve their goals, which could indicate an exemplary display of active inference (Friston et al., 2017b). Given that satisficers exhibited strong differences in RT from all other groups within the first minute of the task, it is likely that satisficers did pursue a different strategy from the outset. The current task is simply unable to identify how effective the satisficers were at adapting to new task demands. The next version of the task will trigger a rule change when the satisficing approach is taken, thereby testing the ability of a satisficer to either shift their strategy to that of an adapter or to seek a new probabilistic regularity. Additionally, increasing the cost incurred for bad decisions across iterations may force a shift in strategy.

If there are trait-level differences between the two successful strategies, then it is important to identify and define those differences. The current labels for these strategies are descriptive. Adapters were labelled “adapters” because they adapted to a new rule, while satisficers were termed “satisficers” because they identified an imperfect but good-enough approach—“satisficing” is a term borrowed from the problem-solving literature which refers to the use of good-enough imperfect solutions (Simon, 1967). However, while writing this paper, we identified a highly relevant field of decision-making which examines the tension between satisficing and *maximizing*, the latter of which describes a problem-solving strategy to identify the maximally perfect solution (Shortland et al., 2018). We propose that the DOT task may assess trait-level maximizing or satisficing, but because we did not include validated maximizing measures (Shortland et al., 2020; Turner et al., 2012), we cannot make any conclusions yet. We plan to include those scales in the next phase of this research to test if the DOT task is effective at predicting trait-level satisficing or maximizing. Given the relevance of satisficing and maximizing in military decision making (Parker et al., 2007; Shortland et al., 2020), a behavioral assessment for such traits may be highly useful for assessment and training.

The next validity check will be comparing performance in the DOT task to performance in the tasks which informed the DOT task: the WCST and the IGT. In the next phase, participants will complete a version of the DOT task as well as the WCST and the IGT. These tasks do not allow for multiple strategies, as there is no probabilistic regularity in the WCST and there is no way to achieve complete certainty in the IGT. If the DOT task identifies trait-level problem-solving approaches, then it is possible that adapters will perform more effectively than satisficers at the WCST, whereas the satisficers will perform more effectively than the adapters at the IGT. Such a finding would provide additional evidence that the DOT task can assess differential levels of sensitivity to prediction error because the two foundational tasks differ in how much uncertainty can be resolved. While adapters may be highly able to rapidly revise their mental models in the WCST, they may be overly sensitive to random losses in the IGT and may erroneously revise their mental models in pursuit of impossible perfection. Comparatively, the satisficers may be able to maintain an effective strategy during the IGT even in the face of random losses but may be less adept at building an ideal model during the WCST.

Although there is much validation work to do before we can propose the DOT task is a valid assessment of active inference, we do want to highlight intriguing similarities to another active inference research project by Marković and colleagues (2019). In that research, the participants performed a probabilistic reversal learning task, wherein two choices had different probabilities of gains or losses, and to succeed, participants needed to learn that one choice was more advantageous than the other. However, at specific intervals, the probabilities reversed and participants had to adapt their choices to continue succeeding, very much like an adaptive form of the IGT. The interesting finding we want to address is that 25% of their participants learned to anticipate the timing intervals for when the probability reversal would occur and began probing the alternative even before the probabilities were reversed. Essentially, these participants had learned a latent rule that governed the probability reversals, and the proportion of participants that learned this latent rule was virtually identical to the proportion of participants that learned and adapted to the latent rules governing the DOT task. If that proportion is continually observed in future versions, it is plausible that we have identified a stable individual difference that informs how a particular proportion of the human population approaches learning novel tasks.

In addition to validating our interpretation of the strategic approaches taken by the adapters and satisficers, we also want to identify psychological constructs that predict the ability and/or the approaches taken to succeed at this task. We initially included the MAAS because some research suggested that active inference was predicted by mindfulness attention (Paulus et al., 2019). However, it is possible that the MAAS is not a sufficiently precise assessment of attentional constructs, as has been suggested elsewhere (Grossman, 2011). For future research, we plan to consider more performance-based cognitive assessments (Anderson & Farb, 2020; Was et al., 2011; Cokely et al., 2012) as well as a survey on emotion regulation (Garnefski & Kraaij, 2006) to conduct a broad assessment of the psychological constructs that inform individual capacity to learn and adapt to novel environments. In addition to improving the study design to examine construct and criterion validity, we will also incorporate more complex statistical analyses to examine participants' process in developing their strategies, such as piecewise regression (Devaraj & Jiang, 2019) or machine-learning (Moss et al., 2022) to model participants' process of identifying and adapting a problem-solving strategy over time.

Army Application

The Army needs the ability to rapidly assess fitness for military service (Yerkes, 1941), and Army psychologists have spent a century assessing general intelligence, job aptitudes, and other individual differences to enable this goal (Goodwin, 2015; National Research Council, 2015). If the DOT task does indeed predict skillful decision making under uncertainty while also diagnosing efficacy to apply specific strategies, all while requiring 15-20 minutes, then the DOT task could provide great value to Army assessment procedures. Military decision making is many things, but it is rarely clean and predictable. Sometimes, effective decision making is about selecting the "least-worst" decision instead of the "perfect" decision (Shortland et al., 2020). Consider General Patton's perspective (1947[1983]) described in *War as I Knew It*, "A good plan violently executed *Now*, is better than a perfect plan next week." However, other military problems are more complex than they initially appear to be and require intense study over time (Grome et al., 2020). Managing complex problems often requires analysis and a continual probing of the environment to learn how the problems respond to different actions and solutions.

Army officers find themselves needing to handle both aspects of problem complexity during their career. The current research developed a prototype task that may be able to assess not only a generalized efficacy to solve novel problems, but also efficacies at applying specific strategies. With more research on the utility and appropriateness of these strategies to real world performance, the DOT task may be further refined as a predictor of a leader's fitness for operational decision making.

The DOT task also has relevance to current Army leadership doctrine; including the Army's Leadership Requirements Model (LRM; Department of the Army, 2022), specifically the Intellect attribute of the LRM. The Intellect requirement describes Army officers' mental agility, as well as their ability to innovate, make sound judgments, and develop expertise. All these requirements appear necessary for exemplary performance in the DOT task. The DOT task also relates to the Develops requirement of the LRM. To develop any skill-based competency, an officer needs to be able to identify and make sense of tasks that will require the target competency. An officer who can rapidly ascertain the demands of a task will be more efficient in managing and accomplishing the task. We believe the DOT task may especially assess generalized tendencies for making sense of novel situations, meaning that performance on the DOT task may predict performance across a variety of contexts in which officers lack prior knowledge and experience. Overall, by continuing this research effort, which will include validating the DOT task and exploring the viability of feedback and improvement, we can provide the Army with a doctrinally framed tool to identify Army officers' capability for making sense of, deciding, and acting in dynamic and uncertain environments.

Finally, we return to our original discussion of the OODA loop. In conversations with PME instructors who have taught the OODA loop, we noted that they often report the most misunderstood component of the model is Orient; specifically, students struggle to understand that orientation is not referring to simplistic ideas like *focusing on* or *attending to* something. Orientation is an active process of binding disparate information about the world into an understanding—that is, a mental model. Essentially, orientation creates a perspective that shapes how an observer understands their environment. We argue that the DOT task is a promising empirical assessment and training product for the OODA loop. Successful performance in the DOT task can be achieved through two distinct successful orientations, satisficing vs. adapting, which could effectively be used to train the process of going through an OODA loop. With two distinct approaches to orientation, which are clearly delineated through behavioral patterns, the DOT task could be a valid empirical assessment that provides a quantifiable indicator of success in achieving as well as adapting the orientation by moving through an OODA loop.

Conclusion

Army officers will always find themselves in unexpected situations to which they must adapt. Whether the problems they face are simple or complex, Army officers will need to make well-reasoned and justifiable decisions based on incomplete information and minimal prior experience. We propose that the competency of active inference is pertinent to Army officers' capacity to make sense of problems they have never encountered before, and that the Army would benefit from a validated measure to assess active inference. The current research developed a prototype measure for assessing active inference ability. We plan to continue this research to hone the efficacy of the DOT task for assessing Army officers' capability for navigating and making sense of dynamic and uncertain situations. To quote the Prussian general von Clausewitz from his treatise *On War*:

> War is the realm of uncertainty; three quarters of the factors on which action in war is based are wrapped in a fog of greater or lesser uncertainty. A sensitive and discriminating judgment is called for; a skilled intelligence to scent out the truth. (von Clausewitz, 1976, p. 101)

Active inference less poetically describes a concept like Clausewitz's "skilled intelligence to scent out the truth." We plan to continue this research to better assess active inference and identify its mechanisms to learn how it operates in exemplary practitioners. With a means to assess active inference capability in Army officers, the Army will be able to create and evaluate developmental interventions to enhance this competency—a competency we believe is critical to officers being able to "scent out the truth," and to do so in complex and uncertain situations which, despite presenting only incomplete and ambiguous information, still demand a decision.

References

Anderson, T., & Farb, N. A. S. (2020). The Metronome Counting Task for measuring meta-awareness. *Behavior Research Methods, 52*(6), 2646-2656. https://doi.org/10.3758/s13428-020-01418-z

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390-412. https://doi.org/10.1016/j.jml.2007.12.005

Barceló, F. (2021). A predictive processing account of card sorting: Fast proactive and reactive frontoparietal cortical dynamics during inference and learning of perceptual categories. *Journal of Cognitive Neuroscience*, *33*(9), 1636–1656. https://doi.org/10.1162/jocn_a_01662

Bartholomees Jr, J. B. (2010). *The U.S. Army War College Guide to National Security Issues, Vol. 1: Theory of War and Strategy*. U.S. Army War College Strategic Studies Institute. https://apps.dtic.mil/sti/pdfs/ADA564450.pdf

Bates, D. (2010). lme4: Linear mixed-effects models using S4 classes. *http://CRAN. R-project. org/package= lme4*.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., & White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*(3), 127-135. https://doi.org/10.1016/j.tree.2008.10.008

Boyd, J. R. (1996). *The essence of winning and losing*. Retrieved from http://www.d-n-i.net/fcs/ppt/boyds_ooda_loop.ppt

Bressan, P. (2002). The connection between random sequences, everyday coincidences, and belief in the paranormal. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 16*(1), 17–34. https://doi.org/10.1002/acp.754

Brown, K. W. & Ryan, R. M. (2003). The benefits of being present: Mindfulness and its role in psychological well-being. *Journal of Personality and Social Psychology, 84*(2), 822–848. https://doi.org/10.1037/0022-3514.84.4.822

Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. *Neuropsychology Review*, *19*(1), 102-114. https://doi.org/10.1007/s11065-009-9083-4

Chan, R. C., Shum, D., Toulopoulou, T., & Chen, E. Y. (2008). Assessment of executive functions: Review of instruments and identification of critical issues. *Archives of Clinical Neuropsychology*, *23*(2), 201-216. https://doi.org/10.1016/j.acn.2007.08.010

Chelune, G. J., & Baer, R. A. (1986). Developmental norms for the Wisconsin Card Sorting test. *Journal of Clinical and Experimental Neuropsychology*, *8*(3), 219–228. https://doi.org/10.1080/01688638608401314

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36*(3), 181–204. https://doi.org/10.1017/S0140525X12000477

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 335-359. https://doi.org/10.1016/S0022-5371(73)80014-3

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159. https://doi.org/10.1037/0033-2909.112.1.155

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The berlin numeracy test. *Judgment and Decision Making*, *7*(1), 25–47. https://doi.org/10.1017/S1930297500001819

Corlett, P. R., & Fletcher, P. C. (2012). The neurobiology of schizotypy: Fronto-striatal prediction error signal correlates with delusion-like beliefs in healthy people. *Neuropsychologia*, *50*(14), 3612–3620. https://doi.org/10.1016/j.neuropsychologia.2012.09.045

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current Directions in Psychological Science*, *19*(1), 51–57. https://doi.org/10.1177/0963721409359277

Damasio, A. R. (1996). The somatic marker hypothesis and the possible functions of the prefrontal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *351*(1346), 1413-1420. https://doi.org/10.1098/rstb.1996.0125

Deane, G., Miller, M., & Wilkinson, S. (2020). Losing Ourselves: Active Inference, Depersonalization, and Meditation. *Frontiers in Psychology*, *11*(539729), 1-15. https://doi.org/10.3389/fpsyg.2020.539726

Deary, I. J., Liewald, D., & Nissan, J. (2011). A free, easy-to-use, computer-based simple and four-choice reaction time programme: the Deary-Liewald reaction time task. *Behavior Research Methods*, *43*, 258-268. https://doi.org/10.3758/s13428-010-0024-1

Dempsey, M. (2012). Mission command white paper. *Department of Defense*. https://www.jcs.mil/Portals/36/Documents/Publications/missioncommandwhitepaper2012.pdf

Department of the Army (2022). *Developing Leaders* (FM 6-22). Author. https://armypubs.army.mil/epubs/DR_pubs/DR_a/ARN36735-FM_6-22-000-WEB-1.pdf

Devaraj, S., & Jiang, K. (2019). It's about time–A longitudinal adaptation model of high-performance work teams. *Journal of Applied Psychology*, *104*(3), 433-447. https://doi.org/10.1037/apl0000372

Diaz-Asper, C., Schretlen, D., & Pearlson, G. (2004). How well does IQ predict neuropsychological test performance in normal adults? *Journal of the International Neuropsychological Society 10*(1), 82–90. https://doi.org/10.1017/S1355617704101100

Euler, M. J. (2018). Intelligence and uncertainty: Implications of hierarchical predictive processing for the neuroscience of cognitive ability. *Neuroscience & Biobehavioral Reviews*, 94, 93–112. https://doi.org/10.1016/j.neubiorev.2018.08.013

Fiorillo, C. D. (2010). A neurocentric approach to Bayesian inference. *Nature Reviews Neuroscience, 11*(8), 605-605. https://doi.org/10.1038/nrn2787-c1

Friston, K. (2005). A theory of cortical responses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *360*(1456), 815-836. https://doi.org/10.1098/rstb.2005.1622

Friston, K. (2008). Hierarchical models in the brain. *PLoS Computational Biology*, *4*(11), e1000211. https://doi.org/10.1371/journal.pcbi.1000211

Friston, K. (2009). The free-energy principle: A rough guide to the brain? *Trends in Cognitive Sciences*, *13*(7), 293-301. https://doi.org/10.1016/j.tics.2009.04.005

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, *11*(2), 127-138. https://doi.org/10.1038/nrn2787

Friston, K. J., Lin, M., Frith, C. D., Pezzulo, G., Hobson, J. A., & Ondobaka, S. (2017a). Active inference, curiosity and insight. *Neural Computation*, *29*(10), 2633–2683. https://doi.org/10.1162/neco_a_00999

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2017b). Active inference: A process theory. *Neural Computation*, *29*(1), 1-49. https://doi.org/10.1162/NECO_a_00912

Garnefski, N., & Kraaij, V. (2006). Relationships between cognitive emotion regulation strategies and depressive symptoms: A comparative study of five specific samples. *Personality and Individual Differences, 40*(8), 1659–1669. https://doi.org/10.1016/j.paid.2005.12.009

Goodwin, G. F. (2015). *U.S. Army Research Institute for the Behavioral and Social Sciences 1940-2015: 75 years of Science and Innovation* (Special Report 75). U.S. Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/citations/AD1007292

Grant, D. A., & Berg, E. (1948). A behavioral analysis of degree of reinforcement and ease of shifting to new responses in a Weigl-type card-sorting problem. *Journal of Experimental Psychology*, *38*(4), 404–411. https://doi.org/10.1037/h0059831

Graves, T. R., Blankenbeckler, P. N., & Stallings, G. M. (2017). *Enhancing fire control decision-making with the Patriot Cognitive Skills Trainer: Development and validation* (Research Report 2011). U.S. Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/pdfs/AD1044544.pdf

Gosling, S. D., Rentfrow, P. J., & Swann Jr., W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Grome, A., Weyhrauch, W. S., Crandall, B., Polander, E., & Laufersweiler, D. (2020). *Managing complex problems: A synthesis of research on Army Design Methodology and strategic thinking* (Research Report 2029). U.S. Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/citations/AD1109776

Grossman, P. (2011). Defining mindfulness by how poorly I think I pay attention during everyday awareness and other intractable problems for psychology's (re)invention of mindfulness: Comment on Brown et al. (2011). *Psychological Assessment, 23*(4), 1034–1040. https://doi.org/10.1037/a0022713

Hamari, J., Koivisto, J., & Sarsa, H. (2014). *Does gamification work?—A literature review of empirical studies on gamification*. In proceedings of the 47th Hawaii International Conference on System Sciences, Hawaii, USA, January 6-9, 2014. https://doi.org/10.1109/HICSS.2014.377

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52*, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679-709. https://doi.org/10.1037/0033-295X.109.4.679

Hutchinson, J. B., & Barrett, L. F. (2019). The power of predictions: An emerging paradigm for psychological research. *Current Directions in Psychological Science, 28*(3), 280-291. https://doi.org/10.1177/0963721419831992

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237-251. https://doi.org/10.1037/h0034747

Kopp, M., Gröblinger, O., & Adams, S. (2019). Five common assumptions that prevent digital transformation at higher education institutions. *INTED2019 Proceedings, 13th International Technology, Education and Development Conference* (pp. 1448-1457). https://doi.org/10.21125/inted.2019.0445

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lewin, K. (1943). Defining the field at a given time. *Psychological Review, 50*(3), 292-310. https://doi.org/10.1037/h0062738

Linson, A., Clark, A., Ramamoorthy, S., & Friston, K. (2018). The active inference approach to ecological perception: General information dynamics for natural and artificial embodied cognition. *Frontiers in Robotics and AI, 5*(21), 1-22. https://doi.org/10.3389/frobt.2018.00021

Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, *49*, 1494-1502. https://doi.org/10.3758/s13428-016-0809-y

Maclean, N. (1992). *Young men and fire*. University of Chicago Press. https://doi.org/10.7208/chicago/9780226501031.001.0001

Manski, C. F. (2017). Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment. *Theory and Decision*, *83*, 155-173. https://doi.org/10.1007/s11238-017-9592-1

Marković, D., Reiter A. M., & Kiebel, S. J. (2019). Predicting change: Approximate inference under explicit representation of temporal structure in changing environments. *PLoS Computational Biology, 15*(1), e1006707. https://doi.org/10.1371/journal.pcbi.1006707

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology*, *41*(1), 49-100. https://doi.org/10.1006/cogp.1999.0734

Moss, J., Wong, A. Y., Durriseau, J. A., & Bradshaw, G. L. (2022). Tracking strategy changes using machine learning classifiers. *Behavior Research Methods*, *54*, 1818-1840. https://doi.org/10.3758/s13428-021-01720-4

National Research Council. (2015). *Measuring human capabilities: An agenda for basic research on the assessment of individual and group performance potential for military accession.* National Academies Press. https://apps.dtic.mil/sti/pdfs/AD1075310.pdf

Patton, G. (1947[1983]). *War as I knew it: The battle memoirs of "blood n' guts."* Bantam.

Parker, A. M., De Bruin, W. B., & Fischhoff, B. (2007). Maximizers versus satisficers: Decision-making styles, competence, and outcomes. *Judgment and Decision Making*, *2*(6), 342. https://doi.org/10.1017/S1930297500000486

Paulus, M. P., Feinstein, J. S., & Khalsa, S. S. (2019). An active inference approach to interoceptive psychopathology. *Annual Review of Clinical Psychology*, *15*(1), 97–122. https://doi.org/10.1146/annurev-clinpsy-050718-095617

R Core Team. (2013). R: A language and environment for statistical computing. Available online, at: http://www.R-project.org/

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64–99). Appleton Century Crofts.

Rominger, C., Fink, A., Weiss, E. M., Schulter, G., Perchtold, C. M., & Papousek, I. (2019). The propensity to perceive meaningful coincidences is associated with increased posterior alpha power during retention of information in a modified Sternberg paradigm. *Consciousness and Cognition*, *76*, 102832. https://doi.org/10.1016/j.concog.2019.102832

Rominger, C., Weiss, E. M., Fink, A., Schulter, G., & Papousek, I. (2011). Allusive thinking (cognitive looseness) and the propensity to perceive “meaningful” coincidences. *Personality and Individual Differences, 51*(8), 1002-1006. https://doi.org/10.1016/j.paid.2011.08.012

Shadrick, S. B., Crabb, B. T., Lussier, J. W., & Burke, T. J. (2007). *Positive transfer of adaptive battlefield thinking skills* (Research Report 1873). U.S. Army Research Institute for the Behavioral and Social Sciences. https://apps.dtic.mil/sti/pdfs/ADA470215.pdf

Shortland, N. D., Alison, L. J., & Moran, J. M. (2019). *Conflict: How soldiers make impossible decisions.* Oxford University Press. https://doi.org/10.1093/oso/9780190623449.001.0001

Shortland, N., Alison, L., & Thompson, L. (2020). Military maximizers: Examining the effect of individual differences in maximization on military decision-making. *Personality and Individual Differences*, *163*, 110051. https://doi.org/10.1016/j.paid.2020.110051

Shortland, N., Alison, L., & Barrett-Pink, C. (2018). Military (in)decision-making process: a psychological framework to examine decision inertia in military operations. *Theoretical Issues in Ergonomics Science, 19*(6), 752-772. https://doi.org/10.1080/1463922X.2018.1497726

Simon, H. A. (1967). Motivational and emotional controls of cognition. *Psychological Review, 74*(1), 29–39. https://doi.org/10.1037/h0024127

Stothart, C., Normand, S., Burbelo, G., & Young, R. (2023). *Thinking-in-time: A scenario-based developmental method for Army officers*. Conference presentation at U.S. Army War College Strategic Leadership Forum, Army Heritage and Education Center, Carlisle, PA, 25 January 2023. https://www.youtube.com/watch?v=4lQlRjtj84c

Sutton, R. S., & Barto, A. G. (1981). Toward a modern theory of adaptive networks: Expectation and prediction. *Psychological Review*, *88*(2), 135-170. https://doi.org/10.1037/0033-295X.88.2.135

Thorndike, E. L. (1927). The law of effect. *The American Journal of Psychology*, *39*(1/4), 212-222. https://doi.org/10.2307/1415413

Toplak, M. E., Sorge, G. B., Benoit, A., West, R. F., & Stanovich, K. E. (2010). Decision-making and cognitive abilities: A review of associations between Iowa Gambling Task performance, executive functions, and intelligence. *Clinical Psychology Review*, *30*(5), 562-581. https://doi.org/10.1016/j.cpr.2010.04.002

Turner, B. M., Rim, H. B., Betz, N. E., & Nygren, T. E. (2012). The maximization inventory. *Judgment and Decision Making*, *7*(1), 48–60. https://doi.org/10.1017/S1930297500001820

van Elk, M., Friston, K., & Bekkering, H. (2016). The experience of coincidence: An integrated psychological and neurocognitive perspective. In K. Landsman & E. van Wolde (Eds.), *The challenge of chance* (pp. 171–185). Springer. https://doi.org/10.1007/978-3-319-26300-7_9

von Clausewitz, C. (1832[1976]). *On war*. Princeton University Press.

Was, C. A., Rawson, K. A., Bailey, H., & Dunlosky, J. (2011). Content-embedded tasks beat complex span for predicting comprehension. *Behavior Research Methods, 43*, 910-915. https://doi.org/10.3758/s13428-011-0112-x

Welsh, M. C., Pennington, B. F., & Groisser, D. B. (1991). A normative-developmental study of executive function: A window on prefrontal function in children. *Developmental Neuropsychology*, *7*(2), 131–149. https://doi.org/10.1080/87565649109540483

Yerkes, R. M. (1941). Psychology and defense. *Proceedings of the American Philosophical Society*, *84*(4), 527-542.

Appendix. Background Survey

The information below is for research purposes only. Your information will not be reported at the individual level. No personally identifiable information will be included in any report or documents resulting from this research.

Current Rank:	________________	Years of Service:	________________
Current MOS/AOC:	________________	Age (in years):	________________

Below, please list any hobbies you have now or have had in the past, with the first being the hobby you have done the longest, the second being the hobby you have done the second longest, etc. Please list no more than three hobbies.

Hobby	Number of Years You Did the Hobby

If you were forced to take a month of annual leave starting one week from now, what would you do during that month? Below, please list the top three activities you would do, with the first activity being the one you would spend the most amount of time on, the second activity being the one you would spend the second most amount of time on, and the third activity being the one you would spend the third most amount of time on.

Activity

Below, please list three positions you have held in the Army (for example, platoon leader, squad leader, etc.), with Position 1 being the position you have held the longest, Position 2 being the position you have held the second longest, and Position 3 being the position you have held the third longest. Additionally, for each position, please list the three activities you conducted the most while holding the position, with Activity 1 being the activity you did the most often, Activity 2 being the activity you did the second most often, and Activity 3 being the activity you did the third most often.

POSITION 1: ______________________ # Years You Held Position: _____

Activity 1: ______________________

Activity 2: ______________________

Activity 3: ______________________

POSITION 2: ______________________ # Years You Held Position: _____

Activity 1: ______________________

Activity 2: ______________________

Activity 3: ______________________

POSITION 3: ______________________ # Years You Held Position: _____

Activity 1: ______________________

Activity 2: ______________________

Activity 3: ______________________

Below, please list any educational degrees/certificates you have (including from PME), with the first one being your highest degree/certificate earned, the second one being your second highest degree/certificate earned, etc. Please list no more than three degrees/certificates.

	Degree/Certificate Concentration	Degree/Certificate Type
Example:	*History*	*Associate's Degree*
	__________	__________
	__________	__________
	__________	__________

What is an unusual skill or ability that you have developed outside of the Army that helps you do your job? ____________________