IPME

INTERNATIONAL PERSPECTIVES ON MILITARY EDUCATION

Volume 2 | 2025





USING TACTICAL DECISION GAMES AS A COGNITIVE ASSESSMENT AND DEVELOPMENT TOOL

Lieutenant Commander Adam T. Biggs, PhD (USN); and Evan D. Anderson, PhD

Abstract: Tactical decision games have been employed by the military for decades as a method to train and practice tactical thinking. A typical game scenario asks participants to review a tactical situation and draft orders on how to respond. These orders are then examined and debated among experts as to the best courses of action. Performance evaluation is often qualitative rather than quantitative, making structured feedback or scoring difficult and limiting opportunities for development of underlying cognitive abilities. The current review provides oversight of five different scoring protocols that can be deployed to support standardized administration of tactical decision games. These scoring methods include illustration, checkbox, expert matching, cognitive probing, and expert debate. Current applications often use expert debate without following other steps in sequence, potentially leaving novice learners to abduce more general tactical principles from specific tactical solutions. If used in sequence, these scoring procedures may enhance tactical learning and development of relevant cognitive abilities while also providing more valuable performance feedback. By adopting and implementing scoring systems appropriate to the current tactical development level, it is possible to expedite tactical learning and development among military personnel.

Keywords: decision-making, educational games, military, tactical, cognitive development

https://doi.org/10.69977/IPME/2025.001

Decision-making during tactical military situations is challenging and time-sensitive, due to the dynamic, high-risk, and inherently uncertain nature of combat environments. Possible choices and their outcomes are sensitive to everything from weapons and personnel available to the weather conditions. Moreover, an "optimal" choice may be relative if the scenario presents a series of tough decisions, all with adverse outcomes. In such cases, underlying preferences (e.g., risk tolerance, speed versus security) may fundamentally vary between decision-makers. Currently, quantitative methods for measuring decision-making performance under such uncertain conditions remain limited. This shortcoming complicates any performance measurement or cognitive learning opportunity during decision-making, especially if teaching individuals to make the fastest and most accurate decisions under stress. Often, decision-making skills are developed through key exercises such as wargaming. Although highly beneficial in helping people understand the military decision-making process, wargaming and military exercises incur a significant burden in time and effort to execute.[1] This approach limits the "reps and sets" for individuals trying to develop their decision-making process. Thus, there is a need for flexible, rapidly applied tools that can facilitate the learning experience.

One key educational tool currently used to improve decision-making is the tactical decision game (TDG). Military leaders have used these games for centuries to develop subordinate decision-making skills and facilitate contingency planning. For example, Helmuth von Moltke would visit Prussian military academies to present a situation, provide instruction, and guide subsequent discussions after students had engaged in their training games.[2] Essentially, a TDG presents players with information about a hypothetical battlefield or tactical scenario. Participants review the material and then construct an attack plan based on available information. The primary purpose of this exercise is to provide some insight into the decision-making process by evaluating a plan produced under some time pressure. Experts will evaluate and de-

[1] LtCol William J. Cojocar, "Adaptive Leadership in the Military Decision Making Process," *Military Review* 91, no. 6 (2011): 29–35.

[2] Maj Donald E. Vandergriff, "From Swift to Swiss: Tactical Decision Games and Their Place in Military Education and Performance Improvement," *Performance Improvement* 45, no. 2 (2006): 30–39, https://doi.org/10.1002/pfi.2006.4930450207.

bate the merits of different plans, which can provide an effective tool to develop tactical decision prowess among military personnel. However, while these games are clearly relevant to battlefield decisions, they are generally not delivered in a systematic curriculum or scored to quantitatively assess performance. Outcomes are instead evaluated qualitatively, with the debate focusing on the merits and shortcomings of each proposed plan as part of the learning opportunity. As such, TDGs are difficult to standardize and evaluate programmatically, limiting their effectiveness as an educational tool. In this sense, TDGs and wargaming exist in a similar state today as might have been executed with the Prussian military. These exercises can and should be developed more formally to underscore advanced cognitive theory and modern ideas of cognitive development or cognitive enhancement.

The current discussion will explore how TDGs might be employed as quantifiable cognitive assessment tools, and how more structured implementation might allow these exercises to aid in cognitive development and enhancement of decision-making. First, the authors review decision-making, decision games, and their application within a professional military education format. Second, a five-step structure will be provided to progressively increase the complexity of evaluating military decision games and probe the underlying cognitive strategies deployed. If applied appropriately, this progression can also be used as a tool to aid cognitive development in those specific decision-making approaches, as well as facilitate the development of tactical capabilities. Third, the discussion will conclude with how these cognitive development steps could optimize any existing utilization of such games. Taken together, the goal is to formalize the use of TDGs and comparable tools to conduct cognitive assessments while simultaneously enhancing the cognitive capabilities of military personnel.

Decision Games for Military Applications

Naturalistic decision-making emerged during the late 1980s to explore how people make decisions in real-world settings.[3] Previously, research across fields such as game theory and behavioral economics had adopted the perspective that humans make

[3] Gary Klein, "Naturalistic Decision Making," *Human Factors* 50, no. 3 (2008): 456–60, https://doi.org/10.1518/001872008X288385.

optimal choices. When provided with complete information, the rational decision-maker would maximize the expected utility (i.e., value, reward) of a decision's outcome. As an account of real-world decision-making behavior, this approach was fundamentally flawed.[4] Evidence from the heuristics and biases literature instead demonstrated that decision-makers often do not rely on logical or rigorous strategies, and they can have great difficulty arriving at mathematically optimal outcomes.[5] Such findings have motivated wide-ranging literature describing how individuals reason and make decisions in complex real-world environments.

The current naturalistic decision-making literature considers four broad categories of decision-making strategy: recognition-primed, rule-based, analytical, and creative.[6] These four strategies (or often, an overlapping combination of them) broadly describe the cognitive strategies deployed when faced with the capacity-limited, information-limited, and time-limited nature of high-risk decision-making (e.g., in legal, military, healthcare, or emergency response settings). Currently, little research exists to directly compare the efficacy of these strategies in terms of increased performance or improved outcomes. The limited existing research suggests the utilization and efficacy of naturalistic decision-making strategies is highly context-dependent, particularly the ability to apply recognition-primed decision-making when addressing time-sensitive environments.[7]

There is a continuing debate about how these broad cognitive strategies relate to specific facets of decision-making ability. The error-prone, quantitatively "irrational" choices produced by naturalistic decision-makers have been variously argued to be the result of cognitive limitations (neoclassical economics), intuitive bi-

[4] Herbert A. Simon, "A Behavioral Model of Rational Choice," *Quarterly Journal of Economics* 69, no. 1 (1955): 99–118, https://doi.org/10.2307/1884852.
[5] Daniel Ellsberg, "Risk, Ambiguity, and the Savage Axioms," *Quarterly Journal of Economics* 75, no. 4 (1961): 643–69, https://doi.org/10.2307/1884324; and Daniel Kahneman, Paul Slovic, and Amos Tversky, ed., *Judgment under Uncertainty: Heuristics and Biases* (Cambridge, MA: Cambridge University Press, 1982), https://doi.org/10.1017/CBO9780511809477.
[6] Carrie Reale et al., "Decision-Making During High-Risk Events: A Systematic Literature Review," *Journal of Cognitive Engineering and Decision Making* 17, no. 2 (June 2023): 188–212, https://doi.org/10.1177/15553434221147415.
[7] Gary Klein, "Recognition-Primed Decisions," in William B. Rouse, ed., *Advances in Man-Machine Systems Research* (Greenwich, CT: JAI Press, 1989), 47–92; and Reale et al., "Decision-Making During High-Risk Events."

ases that distort belief and choice (the heuristics and biases literature), or to be beneficial and adaptive strategies in the face of uncertainty (ecological rationality).[8] These competing theories of decision-making do not provide clear guidance as to the underlying cognitive processes or operations that constrain naturalistic decision-making performance. This is in part due to a limited understanding of the neurocognitive operations that implement specific decision-making strategies (e.g., generating novel solutions, performing analogical reasoning, applying experience to reason about existing options) under stress and uncertainty, complicating creation of targeted training to enhance decision-making in a military context.

These constraints underscore the challenge of identifying an appropriate set of paradigms and procedures for high-fidelity measurement of decision-making performance in real-world contexts. In essence, research on naturalistic decision-making has struggled with the trade-off between ecological validity and experimental control—a very common problem in applied research—such that, by the time the first formal conference had been held on the topic, at least nine parallel models had already been concurrently developed.[9] While the literature documents these strategies, there is considerably less experimental evidence on their relative performance or underlying neurocognitive differences. One approach to address these problems is therefore to devise a reliable research framework that can measure decision-making performance in an environment structured enough to allow systematic skill learning and cognitive development.

One such research framework with a balance of ecological validity and control is the situational judgment test (SJT), known

[8] Kenneth J. Arrow, "Is Bounded Rationality Unboundedly Rational? Some Ruminations," in Mie Augier and James G. March, eds., *Models of a Man: Essays in Memory of Herbert A. Simon* (Cambridge, MA: MIT Press, 2004), 47–55, https://doi.org/10.7551/mitpress/4709.003.0007; Daniel Kahneman, "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93, no. 5 (December 2003): 1449–75; and Gerd Gigerenzer and Reinhard Selten, "Rethinking Rationality," in Gerd Gigerenzer and Reinhard Selten, eds., Bounded Rationality: The Adaptive Toolbox (Cambridge, MA: MIT Press, 2001), 1–12, https://doi.org/10.7551/mitpress/1654.003.0003.

[9] Raanan Lipshitz, "Converging Themes in the Study of Decision Making in Realistic Settings," in Gary A. Klein, Judith Orasanu, Roberta Calderwood and Caroline E. Zsambok, eds., *Decision Making in Action: Models and Methods* (Norwood, NJ: Ablex Publishing, 1993), 103–37.

in military contexts as a decision game. This technique provides participants with a hypothetical scenario and asks them to select the most appropriate response or rank order from several available responses—critically, without instructing or priming the use of a particular decision strategy. Use of SJTs as a selection tool has been validated among medical personnel, to evaluate professional behavior and as a means to assess nonacademic attributes when evaluating candidates.[10] SJTs can compare favorably with existing neurocognitive tests, while having greater predictive validity for identifying and selecting performance under particular circumstances.[11] These attributes make SJTs, and related decision paradigms, a powerful tool in assessment and selection contexts.

Specific SJTs have been deployed for use among military and first responder personnel.[12] It is common for SJTs to assume the moniker of TDGs when applied to these populations; although, for all intents and purposes, they describe essentially the same basic paradigm. These tasks can fulfill versatile roles in military, medical, and law enforcement training. For example, decision games have been suggested as one technique to help responders prepare for asymmetric threats, to practice emergency management, and to otherwise allow for a cost-effective solution to explore

[10] Elin S. Webster et al., "Situational Judgment Test Validity for Selection: A Systematic Review and Meta-analysis," *Medical Education* 54, no. 10 (2020): 888–902, https://doi.org/10.1111/medu.14201; Sebastian Schubert et al., "A Situational Judgment Test of Professional Behaviour: Development and Validation," *Medical Teacher* 30, no. 5 (2008): 528–33, https://doi.org/10.1080/01421590801952994; Kathryn J. Smith at al., "Development and Validation of a Situational Judgment Test to Assess Professionalism," *American Journal of Pharmaceutical Education* 84, no. 7 (2020), https://doi.org/10.5688/ajpe7771; and Fiona Patterson et al., "Evaluations of Situational Judgment Tests to Assess Non-academic Attributes in Selection," *Medical Education* 46, no. 9 (2012): 850–68, https://doi.org/10.1111/j.1365-2923.2012.04336.x.

[11] Anna Koczwara et al., "Evaluating Cognitive Ability, Knowledge Tests and Situational Judgment Tests for Postgraduate Selection," *Medical Education* 46, no. 4 (2012): 399–408, https://doi.org/10.1111/j.1365-2923.2011.04195.x.

[12] Tonis Männiste, Margus Pedaste, and Roland Schimanski, "Review of Instruments Measuring Decision Making Performance in Military Tactical Level Battle Situation Context," *Military Psychology* 31, no. 5 (2019): 397–411, https://doi.org/10.1080/08995605.2019.1645538; Tonis Männiste, Margus Pedaste, and Roland Schimanski, "Situational Judgment Test for Measuring Military Tactical Decision-making Skills," *Military Psychology* 31, no. 6 (2019): 462–73, https://psycnet.apa.org/doi/10.1080/08995605.2019.1664366; and Maj John F. Schmitt, USMCR, *Mastering Tactics: A Tactical Decision Games Workbook* (Quantico, VA: Marine Corps Association, 1994).

decision-making under uncertainty.[13] There is no practical limit on the application as decision games could cover anything from squad-level encounters to a full division-level military engagement. As the scenario merely provides symbols to represent unit size, there are few limits to the scale or decision-making context to be made for a given scenario. Thus, decision games can easily be adapted to fulfill the tactical needs of a particular population and scaled accordingly without incurring a significant cost.

Despite the potential value in using decision games to evaluate personnel, there are several caveats that can limit their practical value. Primarily, any SJT is inherently limited by the answer set provided to the candidate. Not even experts may universally agree on the optimal course of action for a given situation, so forcing a single choice in such a scenario creates a level of subjectivity around the best answer.[14] If a participant would not engage in a course of action provided by the available options, and open-ended answers would not be a practical solution for evaluators, then the test may not adequately reflect individual thinking or creativity. The problem thus becomes a limitation of assessment via hypothetical scenario, both in the branching depth and the quality of choices provided as potential decisions. Another concern is that SJTs may be better at finding consensus among the actions that should not be taken in a given situation rather than actions that should be taken in the same situation.[15] This issue may also be related to poor quality in constructing suitable answer banks for a given question. Immersion is another potential issue that can impact design and limit learning. Scenarios may be brief in part due to logistical considerations of

[13] Neil R. Hintze, "First Responder Problem Solving and Decision Making in Today's Asymmetrical Environment" (thesis, Naval Postgraduate School, Monterey, CA, 2008); M. Crichton and R. Flin, "Training for Emergency Management: Tactical Decision Games," *Journal of Hazardous Materials* 88, nos. 2–3 (2001): 255–66, https://doi.org/10.1016/S0304-3894(01)00270-9; Margaret T. Crichton, Rhona Flin, and William A. R. Rattray, "Training Decision Makers–Tactical Decision Games," *Journal of Contingencies and Crisis Management* 8, no. 4 (2000): 208–17, https://doi.org/10.1111/1468-5973.00141; and Lawrence G. Shattuck et al., "Tactical Decision Making under Conditions of Uncertainty: An Empirical Study," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 53, no. 4, (2009): 242–46, https://doi.org/10.1177/154193120905300417.

[14] See also Gary Klein, "Performing a project premortem," *Harvard Business Review* 85, no. 9 (2007): 18–19.

[15] Wendy E. de Leng et al., "Integrity Situational Judgment Test for Medical School Selection: Judging 'What to Do' versus 'What Not to Do'," *Medical Education* 52, no. 4 (2018): 427–37, https://doi.org/10.1111/medu.13498.

measuring at scale, which can limit how immersed an individual becomes in any particular hypothetical situation and impair the transfer of any acquired skills or knowledge when deployed in stressful or time-sensitive environments. This limitation primarily affects roleplayed or pen-and-paper games; however, computer-based interfaces or high-fidelity environmental simulations also introduce challenges when creating immersion. For example, introducing avatars within a virtual environment brings the complication of how to properly convey emotions or urgency during the briefing, possibly defraying transfer of learning to real-world situations.[16]

Thus, a significant challenge in utilizing decision games involves finding balance between scenario complexity and options among the answers. Scenario complexity must strike an effective balance between detail, possibilities, and immersion to produce a suitable tool that does not create a logistical burden. Although many factors impact the development of decision games, the predominant factor should always be the expertise and development of the personnel to whom it is applied. For example, simulating battalion-level tactical decisions is likely inappropriate for enlisted military students learning squad tactics for the first time. There may be occasional and specific uses to employ scenarios significant beyond the current developmental level of the target personnel, although such applications should be rare and by exception. Similarly, the content knowledge of the intended learners should also be accounted for when developing an assessment. Many symbols or other factors could represent elements of the hypothetical battlefield, yet they should represent factors commensurate with the current learning and goals of the personnel in question.

Likewise, providing possible answers can provide unique character and purpose to any particular game. One approach to assessing performance may be to ensure that personnel are not

[16] Benjamin Goldberg and Janis Cannon-Bowers, "Feedback Source Modality Effects on Training Outcomes in a Serious Game: Pedagogical Agents Make a Difference," *Computers in Human Behavior* 52 (2015): 1–11, https://doi.org/10.1016/j.chb.2015.05.008; Youssef Shiban et al., "The Appearance Effect: Influences of Virtual Agent Features on Performance and Motivation," *Computers in Human Behavior* 49 (2015): 5–11; and Gillian C. Visschedijk et al., "Modelling Human Emotions for Tactical Decision-making Games," *British Journal of Educational Technology* 44, no. 2 (2013): 197–207, https://doi.org/10.1111/j.1467-8535.2012.01286.x.

making clearly wrong decisions by selecting the most inappropriate answers. However, given the subjectivity involved in selecting an optimal choice of action under uncertainty, even having a singular correct answer may be more challenging than it would seem. The number and complexity of options available may then also create further logistical burdens as participants must evaluate each possibility and graders must determine the relative value of each answer. Unfortunately, a key advantage of decision games in this complexity is also a core weakness. That is, decision games are typically unique to a situation. This distinctiveness means that they can be crafted specifically to meet the boutique intents of the evaluators and may therefore be difficult to standardize. Perhaps the best advice when developing decision games for military application is that the entire game, both the scenario vignette and the possible answers, should never be constructed in a vacuum. The specific application, audience, and administration context should be identified prior to developing the game, which act as the primary guidance for all questions that might arise later, such as additional details to include in the vignette or the number and complexity of possible answers.

A Range of Evaluation Options for Tactical Decision Games

Although SJTs and decision games often have similar formats, employing them as cognitive assessment and development tools introduces several advantages for measuring or enhancing naturalistic decision-making. The authors present five progressive approaches to scoring TDGs that layer greater complexity into the practical application of decision games in military education. We define these scoring approaches for the TDG format as: 1) illustration; 2) checkboxes; 3) expert matching; 4) cognitive probing; and 5) expert debate (table 1). These five methods are presented as a continuum from least complex to most complex that can be applied to modify the content and the scoring of a specific scenario. Each additional level introduces a complexity that alters practical application of the decision game, with the intent of advancing the cognitive development of personnel utilizing these tools. Each provides opportunity for administrators or researchers to quantify performance and provides learners the ability to focus learning

Table 1. Overview of the different methods to evaluate tactical decision games (TDGs)

Evaluation method	Scoring technique	Purpose
Illustration	Detail the steps to produce a response in a TDG	Teach new participants how to create a fragmentary order (FRAGO). This method is for instructional purposes only.
Checkbox	Quantify the percentage of essential components included in the response	Ensure that a participant is aware of all the necessary components. This method differs from illustration because the participant generates the answer and receives a percentage score for completion.
Expert matching	Determine how accurately a participant can align their responses with an expert	The scores are compared against expert answers to the same prompts, usually in the form of multiple-choice questions. Participants can see how often their answers align with experts, and ideally, the participants can see logic as to why the expert chose that response over other options.
Cognitive probing	Ask open-ended questions to assess comprehension of a given situation and alternatives	Participants will receive a prompt and then see additional questions that could change the scenario (e.g., What if airpower changed in this scenario?). The intent is to enhance cognitive skills through techniques such as premortem thinking to deepen the participant's ability to think critically about a given scenario.
Expert debate	Experts provide answers to the same prompt, then debate the merits of their responses	Scoring is largely qualitative in this method, but the purpose is to create a structured forum for debating the tactical advantages of different approaches. Answers are neither right nor wrong, though always debatable. Introducing student learners too early to this technique could overwhelm them and limit their learning in these scenarios.

Source: courtesy of the authors.

through a progression of sequential educational goals (see instructional scaffolding).[17]

Illustration

The first evaluation approach has nothing to do with the quality of the tactical actions developed as part of the decision game. Instead, the first evaluation method is intended to simply illustrate the different components of a fragmentary order (FRAGO) developed in response to the hypothetical scenario. Its purpose is to use a sample FRAGO as an example from which the basics of design and organization can be taught to novice students. As such, it could be counterproductive to have a novice learner draft a FRAGO without first giving them the proper pieces that should enter the tactical plan. This approach has the value of teaching fundamental TDG elements so that learning the format does not interfere with developing the cognitive skills and tactical acumen that the decision game is supposed to foster. Nonetheless, illustration is extremely limited in value as its only real purpose is to explain the decision game tenets to novices. Once an individual understands the format, there is little point in returning to the illustration method except to further introduce new concepts, such as decision games that occur at the battalion level rather than the squad level. Quantitative assessments of the learning that occurs during illustration (e.g., testing identification of the components of a FRAGO) are likely of limited value for assessing decision-making ability.

Illustration is most often used in an educational environment. Students are likely learning the basics of tactics while under formal instruction, and so the illustration technique provides a way to introduce key concepts and formats. An instructor will lay out examples and walk students through fundamentals of tactics for given units, formatting for a given order, or otherwise provide instruction as needed to develop skills. In this sense, illustration does little for formal development or tactical acumen. The purpose is to alleviate a cognitive burden by allowing directed learning of basics so that the individuals are not distracted during later stages of learning. For example, if someone is still learning basic symbols used in the

[17] L. S. Vygotsky, *Mind in Society: The Development of Higher Psychological Processes*, ed. Michael Cole et al. (Cambridge, MA: Harvard University Press, 1978).

games, their decision-making errors may come from a lack of symbols knowledge rather than poor tactical plans or errant situational awareness. Illustration is thus a technique for instruction that helps develop individual comprehension for fundamentals associated with tactical decision-making scenarios.

Checkboxes

The second evaluation approach involves checkboxes. Whereas the first method provided illustrations as comparisons of the different solutions to various decision games, the second method provides an opportunity to ensure completion within a particular game. Again, quality answers are less important than consideration and engagement with the available information for the game. The goal should be to ensure that participants consider all pieces of the tactical puzzle they are presented with. For example, one scenario might provide instructions to take a particular hilltop while ensuring sufficient suppressive fire to secure the safest possible approach. If the FRAGO does not include instructions that appropriately incorporate available artillery into the equation or otherwise ignores critical elements of firepower available during the scenario, then the decision plan would likely be incomplete. Checkboxes are a method to ensure completion and integration of all components that should be represented in any plan.

Moreover, checkboxes provide a quantitative method to evaluate performance. If there are 10 items that the FRAGO should include, then having a checklist of 10 items helps identify individual progress as performance can be quantified and compared across scenarios. Still, the goal is to ensure that a participant addresses all the relevant elements. Optimal execution is a secondary concern, and while some feedback can be given to foster future development, feedback should remain positive and focused on the constructive elements of the plan (as detailed negative feedback would likely detract from the primary purpose of the checkbox approach). Checkbox assessments of this nature will load a broader spectrum of cognitive abilities than the previous illustration method (reflecting larger individual differences in cognition) and may serve as a performance assessment for rule-based decision-making when paired with the instructions to consider all provided information in constructing a FRAGO.

Checkbox methods are best used with students in a learning environment. The goal is to ensure that individuals can include or understand all requisite pieces of the scenario. Without a full understanding of the working product, tactical errors could be attributed to misunderstandings in what should be included rather than poor planning. For example, a checkbox method should evaluate whether an individual has prepared a plan in accordance with all rules and formatting required for the scenario. The evaluation is thus based on completion rather than quality of the assembled plan. Its application ensures that students or other people still learning the base scenario have adequately understood everything that should be included. Once they have developed a sufficient cognitive bandwidth to incorporate all requisite elements, then subsequent steps can focus on refinement and quality. Checkbox techniques represent an advanced instructional tool that can provide objective feedback to students rather than a tool to explore tactical proficiency. From a cognitive perspective, the checkbox method helps identify whether the individual is overwhelmed by the cognitive load imposed by a given situation.

Expert Matching

Expert matching is the third approach to evaluate military decision games, and it is this approach that bears greater resemblance to other SJT evaluation methods. The intent is to have experts provide their answers and then compare the answers of advanced students against the expert evaluations. This method can take several formats, although two approaches dominate. In the first, multiple options are provided, and students must select the best course of action. This approach is relatively straightforward, but it can be supplemented as a teaching tool by having a statement from the expert accompany the answer. This statement provides further insight as to why this option would be better than the other possibilities provided. Alternatively, participants may have to rank order from among the multiple options available. By probing participants for this more nuanced response, it is possible to assay greater variation between individuals and gain further insight into their decision-making processes. Scoring can become more complex as to whether any matches should be counted equally or if matching higher-order options warrants more points, but this approach provides a quantifiable method for matching to expert opinion.

Furthermore, there is an opportunity to execute these ideas in an electronic format to better accommodate large samples. For example, some programs computerize such procedures to both effectively scale learning and produce faster feedback, including the ShadowBox™ training program.[18] Among the options available to evaluate solutions to decision games, expertise matching is perhaps the most widely used and most readily quantified, assuming appropriate expert solutions can be attained, which may prompt further debate among experts given the subjectivity in finding an "optimal" solution to any hypothetical scenario (to say nothing of how uncertainty may preclude such a solution at all). The latter problem can be solved, however, by producing only options that achieve expert consensus before providing such possible answers to student learners.

Expert matching is where tactical comprehension truly begins to develop. Problem solving and other higher-order cognitive functions become fully engaged. Accordingly, the individual is now actively looking for critical information and creating possible solutions to the problem set. If the application remains within guided instruction, these cases are likely advanced students who are almost ready to graduate. A more practical example would be individuals with their gaining unit who continue to learn and develop. From a cognitive perspective, expert matching provides multiple opportunities to engage different cognitive abilities. Guided instruction or feedback can still target cognitive development to specific abilities, which can help an individual develop faster. For example, someone might show greater strengths or weaknesses in sensemaking (e.g., quickly assessing a situation) versus identifying tradeoffs and priorities (e.g., determining among competing priorities in complex situations). Targeted cognitive development would limit the TDGs involved based on individual needs. If someone has a shortfall with sensemaking, then they should focus on this element with initial impressions of the scenario. The problem

[18] Joseph Borders et al., "ShadowBox™: Flexible Training to Impart the Expert Mindset," *Procedia Manufacturing* 3 (2015): 1574–79; Gary Klein and Joseph Borders, "The ShadowBox Approach to Cognitive Skills Training: An Empirical Evaluation," *Journal of Cognitive Engineering and Decision Making* 10, no. 3 (2016): 268–80; and Gary Klein, Neil R. Hintze, and David J. Saab, *Thinking Inside the Box: The ShadowBox Method for Cognitive Skill Development* (Washington, DC: MacroCognition, 2013), 121–24.

becomes trying to assess too much at once. Expert matching is an opportunity to develop cognitive skills, but trying to develop too many simultaneously can be overwhelming and ineffective.

Expert matching also raises an important question: How is expertise defined? TDGs represent a wide variety of scenarios that explicitly attempt to foster debate. Different experts might not agree, and so defining the group of people considered "expert" can influence performance standards. The simplest approach is to allow instructors in professional military education or other formal educational environments to be deemed experts by virtue of their position. This approach is straightforward, but it functions as prima facie logic. Another approach might be to identify people with experience in the field most relevant to the scenario. Artillery personnel can provide input if the scenario calls for fires, whereas military police can identify relevant factors based on their experience. Alternatively, minimal experience might still qualify someone as an expert. Drone warfare is a nascent concept, and some of the foremost experts could be people with experience from the Russia-Ukraine conflict. This example emphasizes how expertise might have different definitions for emerging concepts. Nevertheless, there will not be any singular definition of expert when considering tactics. Positional authority and experience help provide some guidance to defining an expert, although the reality is that there will be some fluid definition of expertise based on the scenario.

Cognitive Probing

The fourth evaluation approach, cognitive probing, represents a significant shift in the purpose of cognitive assessments. Unlike checkboxes or expert matching, cognitive probing is a method to explore intuition, innovation, and decision-making by exploring how fully an individual understands the scenario. The premise is relatively simple. Specifically, cognitive probing begins by identifying a simple set of cognitive principles to be explored further. One excellent application would be to draw on the lessons learned from previous examples of cognitive skills training.[19] This existing research identifies and defines seven core activities that could be

[19] Gary Klein et al., "Cognitive Skills Training: Lessons Learned," *Cognition, Technology & Work* 20, no. 4 (2018): 681–87, https://doi.org/10.1007/s10111-018-0528-5.

enhanced during cognitive skills training. For example, attention management describes how well an individual can recognize and monitor critical information, whereas performing workarounds describes how well an individual considers the situation beyond the current ruleset. Each option provides a way to dive deeper into how well a student truly understands the problem in a decision game. Attention management can be manipulated by exploring what information might be missing that could affect the decision, or performing workarounds can be challenged by exploring how the solution might change with the introduction of a new variable—enemy air support suddenly becoming relevant, as one possible tactical development.

Cognitive probing thus delves deeper into comprehension than expert matching could. With options aligned under expert match, there is no room for growth beyond simply having a more complete match to the expert opinion. With cognitive probing, the intent is to push the boundaries of innovation and explore comprehension by allowing the individual to integrate more complex reasoning and judgment skills without completely breaking the boundaries of the given scenario. These cognitive abilities are likely to underpin the successful deployment of naturalistic decision-making strategies, particularly in novel or variable situations in which existing expertise (and recognition-primed decision-making in general) are insufficient to fully assess the present (novel) problem-state or optimally select between possible choices.

There is no hard limit to the cognitive skills that could be integrated into this approach, providing instructors and researchers a structured tool to prompt and directly elicit specific cognitive facets of decision-making. Gary Klein et al. provide an excellent starting point, but aspects such as anticipating future states could incorporate premortem thinking to further extend cognitive capability in decision-making.[20] Premortem thinking pushes an individual by asking what went wrong with their plan before it goes awry.[21] It represents one of many possible techniques to force an individual to think beyond the next action and into the complexities of how subsequent adversarial actions might affect their tactical plan.

[20] Klein, "Cognitive Skills Training."
[21] Gary Klein, "Performing a Project Premortem," *Harvard Business Review* 85, no. 9 (2007): 18–19.

From an educational perspective, the goal is to truly foster expertise development. Cognitive probing explores how much an individual understands a given scenario and would likely benefit individuals who are transitioning into roles as instructors. Whereas expert matching helped someone isolate particular cognitive skills for development, an individual could still confidently provide a robust answer without fully understanding the context and implications. Cognitive probing explores depth of understanding rather than superficial reasoning. From a cognitive development perspective, cognitive probing fosters creativity by helping an individual explore the boundaries of their current thought processes.

Expert Debate

As the fifth and final technique to evaluate military tactics through decision games, expert debate represents TDG discussion in its truest form—and indeed, the way the games were intended to be played. Using this premise, experts craft solutions and debate the merits of different suggestions. Military professional outlets, such as the Marine Corps Gazette, have published these games for decades as well as different solutions to achieve this precise purpose. Their function is not unlike any other form of expertise development, where truly innovative and expert level performance is developed further through peer review and discussion. However, expert debate should not be the only option. Asking novices to participate among experts is both unfair and likely counterproductive to development. Expert debate is also very difficult to quantify when conducting evaluations. It might be possible to have experts provide grading structures on Likert-type or comparable scales, yet the quality of expert solutions is likely to exceed any meaningful differences if using checkboxes or expert matching. Still, there might be additional value to cognitive probing among expert opinions as a means to provide structure and guidance to any debate, particularly if it is targeted at pursuing specific cognitive abilities or naturalistic decision-making strategies. As method of assessment and instruction though, expert debate represents the far end of the spectrum that takes significant time to achieve, though it affords the ability to elicit expert insight into successful naturalistic decision-making.

When these different assessment techniques are applied at the appropriate level, they can give a snapshot of learning that provides instructors insight into the current cognitive capabilities of their personnel. Checkboxes and expert matching are easily the most reliably quantifiable of the methods, which may explain their predominance in previous uses. Other methods have more qualitative value, although it is possible to assign a structured grading system to qualitative methods to gauge progress. For example, premortem thinking questions can be posed to participants with a rubric to delineate unsatisfactory answers from average or excellent answers. The problem becomes standardization. A relatively objective rubric can provide structure to opinions, but this method becomes difficult to scale as it does require someone to provide feedback to conduct an evaluation. Expert matching, by comparison, scales more easily given that the matching function can be automated and placed into electronic format. Each method has its advantages and disadvantages, both in scalability and the quality of feedback the technique can provide, but there are subsequent advantages to having a more complex and multilayered scoring system to evaluate military decision games.

Enhancing Cognitive Development through Decision Games

The cognitive assessment methods listed previously provide multiple pathways to evaluate decision games for military application. Each method evolves and expands the complexity of scoring, yet this increased complexity comes with a particular purpose. That is, the scoring system evolves along with student capabilities. This progressive instructional approach affords an ideal environment to train cognitive skills. Novice students should be learning the basics of the paradigm and ensuring they can provide a complete order that addresses all components of the scenario. This goal is accomplished by the illustrative nature of the first approach and the checklists of the second approach. Expert matching and cognitive probing further expand these capabilities—first by matching the responses to expert solutions, then by expanding reasoning and decision-making elements within individual comprehension of the scenario. The goal is to reach the final stage, expert debate, on a much faster timeline than what would have been probable otherwise.

Granted, there is a major concern using decision games as a cognitive development tool. Medical studies suggest that there is significant value in using SJTs to evaluate nonacademic aspects, such as professionalism, but there is scant evidence that decision games can be used to expand cognitive development. Expanding the scoring system is one possible approach that could both aid learning and quantifying progress, which might be a suitable method to enhance cognitive development without evoking significant debate about near transfer or far transfer of cognitive learning.[22] Still, the lack of empirical evidence raises questions about how best to utilize decision games in support of cognitive development for servicemembers.

Conversely, elements of naturalistic decision-making (e.g., practicing underlying cognitive skills, developing specific decision-making strategies) amenable to education and enhancement are likely to require some amount of simulated (e.g., TDGs) or real-world (e.g., combat deployment) practice to foster meaningful and sustained enhancements. Four primary strategies for naturalistic decision-making have been studied: recognition-primed, rule-based, analytical, and creative. Of these, intuitive/recognition-primed and rule-based strategies are the most likely to benefit from intentional cognitive development through TDGs, as they depend directly on the structured memory and crystallized knowledge that progressive use of TDG scoring allows participants to acquire. In contrast, creative and analytical decision-making may be more difficult to develop or enhance, as they depend on more complex cognitive operations and on resistance to some sources of heuristics and biases. Even so, empirical evidence demonstrates that cognitive flexibility and decision-making are amenable to enhancement through intervention, suggesting that structured education

[22] *Near-transfer* is a term that indicates the training task is close to the actual application, whereas *far-transfer* indicates that the training task is highly different from the end application. For example, a near-transfer marksmanship task would train someone with a pistol on a marksmanship range for a shooting competition. By comparison, a far-transfer task might be cognitive enhancement through a computer-based memory task to help someone handle the cognitive load of an intense real-world shooting scenario. Susan M. Barnett and Stephen J. Ceci, "When and Where Do We Apply What We Learn?: A Taxonomy for Far Transfer," *Psychological Bulletin* 128, no. 4 (2002): 612, https://doi.org/10.1037/0033-2909.128.4.612.

may be capable of driving improvements in performance.[23] Regardless, it indeed remains an empirical question as to the extent decision-making skills taught using simulated games would show transfer to enhanced decision-making performance in real-world combat operations.

An important consideration is that, while the various cognitive assessments may be limited in evaluating cognitive development, there is a key advantage in that they can capture and quantify individual differences in decision-making performance and development of expertise. In particular, expert matching provides an ideal comparison between unaided individual progress and the progression of cognitive development when aided by directed learning programs. Put another way, is the program effective in developing experts faster than student learning without such support? There is an attractive element in quantifying performance and progress among military personnel that cannot be ignored, especially when considering how program efficacy will be reported to senior leadership.[24] For example, in a hypothetical situation, expert matching provides an easy way to quantify those students who underwent a cognitive development initiative and matched experts 96 percent of the time after only 10 weeks, whereas students who did not undergo similar development matched experts only 25 percent of the time during the same period. An extreme example of hypothetical learning potential, but the contrast displays a single-sentence, bullet point method of conveying outcomes that would appeal to military leadership. This quantifiable method of evaluating progress is an important step in measuring program success for any military endeavor.

Given the relative importance of expertise matching in measuring progress, it is also important to stress the nuanced role that expert matching can play in cognitive development. Specifically, early learning steps should have expert matching in scenarios with

[23] Radwa Khalil, Ben Godde, and Ahmed A. Karim, "The Link Between Creativity, Cognition, and Creative Drives and Underlying Neural Mechanisms," *Frontiers in Neural Circuits* 13, no. 18 (2019), https://doi.org/10.3389/fncir.2019.00018; and Christopher E. Zwilling et al., "Enhanced Decision-making through Multimodal Training," *NPJ Science of Learning* 4, no. 11 (2019), https://doi.org/10.1038/s41539-019-0049-x.

[24] Adam T. Biggs, "How to Enhance Military Research Using Mathematical Psychology," *Journal of Mathematical Psychology* 106 (2022): 102619, https://doi.org/10.1016/j.jmp.2021.102619.

consensus. Novice learners should not attempt to pick apart the intricacies associated with divergent expert opinions that address subtle or even theoretical points in application. Instead, novice learners would benefit most from engaging in decision games where the experts have reached consensus on the best courses of action. As the student progresses, it may become valuable to give them scenarios with diverging expert opinions. Seeing and dissecting the contrast may provide a valuable learning opportunity that aids in their development. For similar reasons, expert matching is likely best done initially as a single-choice option, where students are selecting the best course of action. This approach limits the cognitive workload as students are merely identifying the best course of action and not exploring the intricacies of relative merit among lesser options. Because high cognitive workload can impair learning, the associated cognitive load should be appropriately managed during the learning process.[25] The goal should be to challenge students without overwhelming them. During decision games, rank ordering options represent a more complex cognitive task as nuance among lesser options must be debated and compared. As such, rank ordering options should be incorporated as a method to increase the cognitive workload and help students progress to a more advanced state but not introduced as the first approach when conducting expert matching.

Another consideration involves the type of decision-making that military decision games should aim to develop. Combat scenarios inherently have an element of urgency and danger that cannot truly be paralleled in other forms of decisions. For example, while trauma surgery may have significant pressure as the life of the patient could be at risk, the doctor's life is rarely at stake within a surgical setting. The same cannot be said for combat applications or even military medicine. Additional hazards increase the relative dangers associated with decision-making, and in so doing, significantly increase the environmental pressures. Substantial effort has gone into exploring techniques and methods to increase the real-

[25] F. Javier Lerch, Cleotilde Gonzalez, and Christian Lebiere, "Learning under High Cognitive Workload," in *Proceedings of the Twenty First Annual Conference of the Cognitive Science Society* (East Sussex, UK: Psychology Press, 1999), 302–7; and Ryan McKendrick et al., "Theories and Methods for Labeling Cognitive Workload: Classification and Transfer Learning," *Frontiers in Human Neuroscience* 13, no. 295 (2019), https://doi.org/10.3389/fnhum.2019.00295.

ism of such scenarios.[26] For high-stakes decision-making, the implication is that decisions and planning may not always be conducted under the relatively leisurely pace of SJTs or other decision game formats. Military personnel may need to rely more on intuitive decision-making.[27] Intuitive decision-making may depend heavily on recognition-primed strategies, implicating pattern recognition and crystallized knowledge in performance.[28] Threat assessments and related perceptual processes may thus have a more significant influence on intuitive decision-making than a typical decision game would afford to practice, highlighting the utility of training and measuring fast and efficient rule-based decision-making in more realistic contexts. Therefore, cognitive development must consider the need to enhance decision skills in both controlled formats associated with planning and the more urgent formats likely to better parallel combat decisions.

Summary and Future Directions

Decision games provide an excellent tool to teach tactical think-

[26] Katherine R. Gamble et al., "Different Profiles of Decision Making and Physiology under Varying Levels of Stress in Trained Military Personnel," *International Journal of Psychophysiology* 131 (2018): 73–80, https://doi.org/10.1016/j.ijpsycho.2018.03.017; Debbie Patton, "How Real Is Good Enough?: Assessing Realism of Presence in Simulations and Its Effects on Decision Making," in *Foundations of Augmented Cognition: Advancing Human Performance and Decision-Making through Adaptive Systems* (Cham, Switzerland: Springer, 2014), 245–56, https://doi.org/10.1007/978-3-319-07527-3_23; and Debra Patton and Katherine Gamble, "Physiological Measures of Arousal during Soldier-relevant Tasks Performed in a Simulated Environment," in *Foundations of Augmented Cognition*, 372–82, https://doi.org/10.1007/978-3-319-39955-3_35.

[27] Lisa A. Burke and Monica K. Miller, "Taking the Mystery out of Intuitive Decision Making," *Academy of Management Perspectives* 13, no. 4 (1999): 91–99, https://doi.org/10.5465/ame.1999.2570557; and Katherine H. Hall, "Reviewing Intuitive Decision-making and Uncertainty: The Implications for Medical Education," *Medical Education* 36, no. 3 (2002): 216–24, https://doi.org/10.1046/j.1365-2923.2002.01140.x.

[28] Gary Klein, *Sources of Power: How People Make Decisions* (Cambridge, MA: MIT Press, 1998), https://doi.org/10.7551/mitpress/11307.001.0001; Jennifer K. Phillips, Gary Klein, and Winston R. Sieck, "Expertise in Judgment and Decision Making: A Case for Training Intuitive Decision Skills," in *Blackwell Handbook of Judgment and Decision Making*, ed. Derek J. Koehler and Nigel Harvey (Hoboken, NJ: Blackwell Publishing, an imprint of Wiley, 2004), 297–315, https://doi.org/10.1002/9780470752937.ch15; and Gary Klein, "A Naturalistic Decision Making Perspective on Studying Intuitive Decision Making," *Journal of Applied Research in Memory and Cognition* 4, no. 3 (2015): 164–68, https://doi.org/10.1016/j.jarmac.2015.07.001.

ing in a progressive manner. Although these games have been developed and played among military servicemembers for decades, there remains immense potential in enhancing the process. This potential lies practically in truncating the learning curve so that novices become experts in a shorter time frame and theoretically in enhancing cognitive development of the individual, particularly where instruction is scaffolded to target specific cognitive abilities or decision-making strategies. In summary, the steps laid out here provide distinct opportunities to use the scoring system as feedback to expedite individual development of tactical decision-making, and ultimately to measure and track individual differences in naturalistic decision-making performance.

An important question is whether the context changes the implications for different scenarios. For example, TDGs could be applied in a training environment, such as The Basic School at Marine Corps Base Quantico, Virginia, versus a professional military education environment, such as the U.S. Army's Command and General Staff College at Fort Leavenworth, Kansas. Many different environments use some form of TDGs. Still, the application is likely matched most to the expertise of the personnel and their current cognitive enhancement needs. Basic training environments are more likely to use illustration or checkboxes since they are dealing with junior personnel who are learning the fundamentals. Conversely, professional military education would benefit more from cognitive probing methods because the students are likely to be at a more advanced level of their career. The specific application scenario is therefore not as important as the current experience. Even an operational environment could use all five methods, but the best use case will depend entirely on the current expertise of the target audience.

Standardization is another topic to consider when contrasting these methods across different environments. Some standardization is essential, especially at lower levels, such as illustration, where the common operating picture should be similar across different scenarios. Military doctrine largely addresses this need by standardizing symbols across different scenarios. However, the need for standardization becomes less as the method becomes more developed. Expert debate by its very nature cannot fully be standardized as its purpose is to elicit novel ideas. Perhaps expert matching is the one level most in need of formal standardization. Expert matching requires some process to identify experts, create scenarios, develop answers, and refine feedback. As more expert matching can be developed, greater standardization can impart a more consistent level of instruction for different scenarios and applications. Common consensus on the experts involved is implicit in standardization, but these tools are also very important in individual development. Nevertheless, the takeaway is that standardization needs vary across levels of cognitive analysis with TDGs, where increased standardization is more important at lower levels of analysis.

Other dynamic learning methods could similarly benefit from the cognitive structure supplied here. For example, wargaming is often a complex and dynamic training exercise that helps foster professional development and individual expertise in warfighting. However, wargaming is also notoriously reliant on a highly capable white cell to adjudicate the exercise while trying to ensure some level of realism. Wargaming could benefit from the cognitive structure discussed here by facilitating expert matching, cognitive probing, and expert debate in measured response to the needs of individuals completing the exercise. In practice, this change would have one member of the white cell dedicated to cognitive enhancement components for individual development. An example could involve using cognitive probing to assess whether the individual is fully thinking through the problem and pushing the boundaries of their understanding. In practice, a division-level exercise might impose a target working group as part of the simulation but run by the white cell to ask leading questions about situational awareness and problem solving. This working group could maintain the perceived realism of the exercise while providing a targeted opportunity for cognitive enhancement by using methods outlined in the cognitive probing stage. There are many such possible applications that can be tailored to individual wargames, exercises, and other warfighting tasks within an educational environment. For cognitive enhancement, the goal is to supplement activities in the exercise based on cognitive skills in such a way as to augment the individual learning experience and facilitate expertise development.

Finally, it is important to note that we have not addressed the importance of team-based performance. Effective tactical decisions must still be communicated and coordinated, and while the orders imply the capability to do so, execution remains a concern warranting additional attention. This form of team development remains an essential component of effective military performance.[29] Thus, tactical decision-making development should be supplemented by field exercises to ensure that these skills remain viable under highly realistic settings, and future work may therefore wish to extend the approach and scoring formats advocated here to group or collective scenarios.

About the Authors

LtCdr Adam Biggs is a commissioned officer in the U.S. Navy Medical Service Corps. He received a PhD in cognitive psychology from the University of Notre Dame and completed postdoctoral work at Duke University. His research experience includes studies on cognitive enhancement, lethal force decision-making, resilience, and stress. https://orcid.org/0000-0003-3394-0691

Dr. Evan Anderson is a civilian employee of the U.S. Department of the Air Force. He completed his PhD in neuroscience at the University of Illinois at Urbana-Champaign. His research experience involves studies on cognitive enhancement and the neuroscience of decision-making, intelligence, and recovery from traumatic brain injury. https://orcid.org/0000-0003-4975-499X

[29] Joan H. Johnston et al., "A Team Training Field Research Study: Extending a Theory of Team Development," *Frontiers in Psychology* 10 (2019): 1480, https://doi.org/10.3389/fpsyg.2019.01480.