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

Research Report 2018

The Influence of Expertise and Decision Environment on Collective Hypothesis Generation

Drew A. Leins
Jim Leonard
Applied Research Associates

Gary Masters
Engility Corporation

Christopher L. Vowels
U.S. Army Research Institute

January 2018

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

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved for distribution:

MICHELLE L. ZBYLUT, Ph.D.
Director

Technical review by

Dr. Anthony Scroggins, U.S. Army Research Institute
Dr. Rand Brou, U.S. Army Research Institute

NOTICES

REPORT DOCUMENTATION PAGE

1. REPORT DATE (dd-mm-yy)	2. REPORT TYPE Final	3. DATES COVERED (from. . . to) June 2016 – April 2017

4. TITLE AND SUBTITLE The Influence of Expertise and Decision Environment on Collective Hypothesis Generation	5a. CONTRACT OR GRANT NUMBER W5J9CQ-11-0003-0005
	5b. PROGRAM ELEMENT NUMBER 622785
6. AUTHOR(S) Drew A. Leins, Jim Leonard (Applied Research Associates), Gary Masters (Engility Corporation), and Christopher L. Vowels (U.S. Army Research Institute)	5c. PROJECT NUMBER A790
	5d. TASK NUMBER 215
	5e. WORK UNIT NUMBER
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
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. MONITOR ACRONYM ARI
	11. MONITOR REPORT NUMBER Research Report 2018

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

13. SUPPLEMENTARY NOTES
COR and Subject Matter POC: Dr. Christopher L. Vowels

14. ABSTRACT *(Maximum 200 words)*:

In a previous study, we explored the use of heuristics at the individual and collective level. In the current study, collective decision-making was further investigated by examining the influence of shared (deployment) experience on generating hypotheses. Groups of Soldiers viewed and assessed images depicting different decision environments: operational scenarios (familiar task) or medical triage scenarios (unfamiliar task). After assessing initial threat risk, groups monitored situations as more information was added. When incoming information changed their assessment, Soldiers stopped a scenario and reported a new assessment. We tested whether experience interacted with the order of low- and high-value cues, familiarity of the decision environment and time pressure. Experienced groups were more likely to stop and re-evaluate scenarios at optimal times (when high-value cues were present), but only in familiar scenarios. This suggests these experienced groups were utilizing cue-activated stop rules that were better calibrated to the familiar environments. Time pressure did not induce groups of either experience level to reassess scenarios quicker. We did not find a difference of individual contribution to discussions across the different groups; more clearly defined roles within the experiment might have revealed variable contribution as expected from utilization of a transactive memory system.

15. SUBJECT TERMS

Hypothesis, Hypothesis generation, Training, Heuristic, Collective Performance, Squad

SECURITY CLASSIFICATION OF			19. LIMITATION OF ABSTRACT	20. NUMBER OF PAGES	21. RESPONSIBLE PERSON
16. REPORT Unclassified	17. ABSTRACT Unclassified	18. THIS PAGE Unclassified	Unlimited		Brian T. Crabb 254-288-3833

Research Report 2018

The Influence of Expertise and Decision Environment on Collective Hypothesis Generation

Drew A. Leins
Jim Leonard
Applied Research Associates

Gary Masters
Engility Corporation

Christopher L. Vowels
U.S. Army Research Institute

Fort Hood Research Unit
Brian T. Crabb, Chief

U.S. Army Research Institute for the Behavioral and Social Sciences
6000 6th Street, Bldg 1464, Fort Belvoir, VA 22060-5610

January 2018

Army Project Number 622785A790	**Personnel, Performance and Training Technology**

ACKNOWLEDGEMENT

The authors would like to thank the Soldiers and leaders who took time to participate in this research and provide their insights involving individual and collective decision-making. We would also like to thank the unit leadership for finding time to support the data collections among several competing priorities.

THE INFLUENCE OF EXPERTISE AND DECISION ENVIRONMENT ON COLLECTIVE HYPOTHESIS GENERATION

EXECUTIVE SUMMARY

Research Requirement

Research has examined expert decision-making in various domains, ranging from military to healthcare to sports (Zsambok & Klein, 2014) and often found that it appears to be quicker and less effortful than novice decision-making. The recognition-primed decision (RPD) model suggests that these faster, less effortful judgments stem from experts' ability to recognize cues and engage in pattern matching (Klein, 1993, 1997). Through experience in a domain, experts develop deeper knowledge and a larger repertoire of relevant experiences they can retrieve from long-term memory. Drawing on these experiences, experts can quickly recognize potentially relevant information and generate explanations for a situation. This often looks like intuitive rather than effortful decision-making and is particularly useful in time-pressured situations in which decision makers must find alternative or atypical solutions. Novices in these situations have less knowledge and fewer experiences to draw from, often resulting in slower and more deliberate hypothesis generation. However, as we observed in Leins et al. (In Preparation), working in groups can also slow decision-making and may lessen any advantage conferred by individual expertise. To explore this further in the current experiment, we examined groups with different levels of deployment experience.

Similar to individual expertise development as a result of experience, groups can also develop a collective expertise. Groups may develop both task domain knowledge and group-specific knowledge. Similar to task domain knowledge, group-specific knowledge can lead to more efficient decision-making. One path to making group decision-making more efficient is to engage a transactive memory system (Wegner, 1987). Transactive memory refers to the division of cognitive labor among group members whereby, over time, they develop implicit roles for encoding and recalling particular categories of details. As a squad gains operational experience, members may naturally assume various roles aligning with their individual cognitive capabilities. During subsequent group decision-making, the group may divide the cognitive labor according to those implicit roles. In the following experiment, we sought to explore whether Soldiers with more experience working together would generate hypotheses more efficiently than Soldiers with less experience working together, and to see if that effectiveness transferred across domains or if they were domain specific.

Procedure

We used a 2 x 2 x 2 x 2 mixed design in which relevant deployment experience (experienced vs. inexperienced) was the between-groups factor and cue-order (high-value first vs low-value first), task familiarity (familiar vs. unfamiliar), and time pressure (low vs. high) were within-subjects factors. Participants worked collectively to assess the threat risk posed by various scenarios. We examined 115 Soldiers working in groups of 3-4. Sixty-four participants (14 groups) represented experienced Soldiers, as these Soldiers had relevant deployment experience. The remaining 51 participants did not have relevant deployment experience and therefore represented inexperienced Soldiers.

Participants generated hypotheses for 8 scenarios, half of which presented threat detection (familiar) tasks and half presented medical diagnosis (unfamiliar) tasks. Each scenario included a short description accompanied by an image. The descriptions presented the scenario context and specific decision requirements. A request for a threat assessment of the scenario then followed. After groups entered their assessment and confidence rating, a new cue was added to the image every six seconds until four new cues had been added or until a group member stopped the trial to indicate a change in assessment. Groups then discussed whether they would enter a changed collective assessment. If they chose to enter a new assessment, the scenario ended and a new scenario began. If they chose to learn more about a paused scenario, they could resume the scenario and view more information. Each scenario contained one high-value cue presented in either the second or third serial order position. After completing all scenarios, participants completed two decision-making disposition scales, the Decision-Making Style (Scott & Bruce, 1995) and the Need for Cognitive Closure (Roets & Van Hiel, 2011; for the original scale, see Webster & Kruglanski, 1994).

We scored initial and secondary assessments for threat ratings, reasoning, and confidence level, and calculated response latency, quality of hypothesis timing, and confidence scores across scenarios within each condition. To account for group member contribution, we calculated the proportions of contribution for each participant in each scenario and used these proportion scores to calculate the contribution variance for each group across scenarios and conditions.

Results

Initial threat rating. Overall, groups' mean initial threat rating for scenarios across conditions corresponded to a rating between "some risk" and "moderate risk." Neither experience level nor time pressure influenced initial threat ratings. However, task familiarity influenced initial threat ratings: Groups rated familiar scenarios as presenting a higher initial threat risk than that presented by unfamiliar scenarios.

Individual-group differences. Task familiarity also influenced the extent to which individual group members' initial ratings differed from the initial consensus group rating for each scenario. After deliberating on familiar tasks, group consensus ratings were slightly lower than individual group members' initial ratings, whereas in unfamiliar tasks, group consensus ratings were slightly higher than individual group members' initial ratings. Neither experience nor time pressure influenced this difference.

Group confidence. Overall, groups rated their confidence moderately high (roughly 87/100). Neither experience level nor task familiarity affected group confidence ratings. However, time pressure interacted with experience to influence group confidence: Groups with experience reported higher confidence when under time pressure than when under no time pressure, whereas groups without experience reported lower confidence when under time pressure than when under no time pressure.

Change in group threat rating over time. Overall, groups tended to raise their threat ratings by roughly one unit of risk (e.g., from "some risk" to "moderate risk," or from "2" to "3" on a 5-point scale). Only task familiarity influenced changes in threat ratings over time. Groups' threat ratings increased by more in unfamiliar scenarios than in familiar scenarios.

Quality of timing of changes in assessments. In general, experienced groups stopped scenarios sooner than did inexperienced groups. In addition, across experience levels, groups were more likely to stop scenarios on time than early. Deployment experience influenced the timing of changes in assessments. Groups of inexperienced Soldiers were more likely than experienced groups to stop scenarios late. There was a main effect of cue order on when groups stopped scenarios. Groups stopped early-cue scenarios earlier than they stopped late-cue scenarios. However, when examining only on-time assessment changes, cue order did not affect the quality of timing. Although groups in general viewed fewer images in scenarios in which the high-value cue appeared early, they were still no more likely to make on-time reassessments. Task familiarity interacted with experience to influence the quality of timing, whereby experienced groups exhibited better timing than inexperienced groups during familiar scenarios, but not during unfamiliar scenarios. Overall, time pressure did not influence the timing of reassessments. However, time pressure interacted with experience to influence the timing of reassessments. Whereas experienced groups were no more likely to reassess early, on time, or late when under time pressure, inexperienced groups were more likely to reassess scenarios late versus early or on time.

Examining only on-time assessment changes, a repeated-measures ANOVA using cue order, familiarity, and experience as independent variables revealed a three-way interaction effect on the quality of timing. When the scenario was familiar and the high-value cue appeared late, groups of experienced Soldiers were more likely to stop scenarios at optimal times versus when the scenario was unfamiliar and the high-value cue appeared late. There was no similar effect for experienced Soldiers when the high-value cue appeared early. By contrast, the timing of groups of inexperienced Soldiers did not differ as a function of the interaction of cue order and scenario familiarity.

Supplemental analyses. We also explored whether groups differed regarding (a) how much group members contributed to consensus discussions (contribution variance) and (b) group decision-making traits as measured by group mean scores on two decision-making disposition scales, the Decision-Making Style (DMS) scale and the Need for Cognitive Closure scale (NFCC). Two decision-making styles correlated with timing quality: As groups scored higher in either dependent or avoidant decision-making styles, they tended to stop scenarios later. However, only the avoidant decision-making style approached statistical significance in a regression analysis predicting the timing of reassessments. Scores on neither scale correlated with timing of reassessments.

Regarding group members contribution to assessment discussions, we observed no difference in group contribution variance between experienced and inexperienced groups. Moreover, contribution variance did not correlate with the timing of assessment changes, the number of images viewed in any condition.

Utilization and Dissemination of Findings

Groups of experienced Soldiers were generally more efficient at reassessing threats than were groups of inexperienced Soldiers. That is, experienced groups were more likely than inexperienced groups to stop a scenario at the time a highly threat-relevant cue appeared on screen, when the set of cues was most informative. By contrast, inexperienced groups were more likely to continue evaluating cues after the highly relevant cue appeared and well after the cue set's informativeness peaked. Moreover, experienced groups exhibited better timing when scenarios presented a familiar task versus an unfamiliar task, whereas inexperienced groups were equally inefficient across task familiarity conditions. This suggests that experienced groups were more likely than inexperienced groups to engage cue-activated stop rules that were better calibrated to the familiar decision tasks. When experienced Soldiers perceived highly relevant cues in a familiar domain, those cues passed a threshold of informativeness. Those Soldiers stopped searching for additional information, and they reported a reassessment. By contrast, inexperienced Soldiers often failed to differentiate the relevance of those same cues. Thus, high-value cues did not exceed the threshold at which inexperienced Soldiers would stop searching for information, and consequently, these Soldiers continued searching for and evaluating cues. They performed this way whether the scenario depicted a familiar task or an unfamiliar task. These results align with the notion that domain experience allows decision makers to leverage recognition-based heuristics when assessing domain-specific situations (Klein, 1993, 1997; Shanteau, 1992a, 1992b).

In the current study, we observed mixed effects of time pressure. Overall, groups' proportions of on-time reassessments dropped from 45% of all reassessments under no time pressure to 35% of all reassessments under time pressure. However, upon further examination, inexperienced groups drove this effect in a surprising way: They delayed a majority of their reassessments when under time pressure. In fact, neither experienced nor inexperienced groups disproportionately accelerated their decision-making under time pressure. Thus, we found that groups did not make suboptimal early decisions and therefore may have consistently applied recognition-based heuristics with uniform thresholds of informativeness across pressure and no-pressure conditions. This allowed experienced groups to continue to make a substantial number of on-time assessments, particularly on familiar tasks, but it did not similarly benefit inexperienced groups. Inexperienced groups, by nature, did not have the same experiences or rich memory traces to rely on when evaluating the informativeness of environmental cues and thus did not recognize high-value cues as highly threat relevant, regardless of task familiarity.

Regarding collective decision-making, we examined individual contributions to group consensus discussions, and we found no influence of within-group contribution variance on any outcome. Furthermore, we found no influence of deployment experience on within-group contribution variance. If groups were to demonstrate the use of something like a transactive memory system when deliberating, we might expect to see different levels of contribution across different tasks or domains. This was not apparent in the current study. However, the application of recognition-based heuristics among groups may benefit, albeit subtly, from the effective division of cognitive labor that would characterize the use of transactive memory. If so, experienced groups may have indeed leveraged such a memory system here. Future studies can further explore Soldiers' use of transactive memory systems by carefully manipulating the knowledge possessed by each group member, as well as the knowledge required to solve experimental problems. Thus, different constellations of knowledge within a group should differentially confer advantages to that group across problems.

In general, the findings indicate that when assessing threat risk collectively, groups of Soldiers likely rely on recognition-based heuristics for identifying threat-relevant features in an environment. We found a strong effect of domain familiarity on the efficiency with which groups assessed the threat risk in a decision environment. A familiar context and early access to valuable information allowed groups of Soldiers with domain experience to successfully engage heuristics and register assessments efficiently. These findings can help to develop new research questions and to develop better measures of performance for home station training. It appears deployment experience allows Soldiers to use heuristics more efficiently in threat detection contexts with which they are familiar. And those gains in efficiency can be observed in decision-making. As a result, Army leaders and trainers should be able to develop training exercises that provide targeted learning opportunities similar to those provided by deployment experience. Additionally, these findings could allow trainers to more effectively evaluate Soldier and small-unit performance on decision-making in threat detection. Such training and evaluation will help accelerate skills acquisition and Soldier readiness.

THE INFLUENCE OF EXPERTISE AND DECISION ENVIRONMENT ON COLLECTIVE HYPOTHESIS GENERATION

CONTENTS

LIST OF TABLES

LIST OF FIGURES

THE INFLUENCE OF EXPERTISE AND DECISION ENVIRONMENT ON COLLECTIVE HYPOTHESIS GENERATION

In a pair of previous experiments, we examined whether Soldiers employed heuristics when generating hypotheses to explain threats in operational environments (Leins, Leonard, Zimmerman, Minchew, & Vowels, (In Preparation). Specifically, we explored whether Soldiers employed cue-activated stop rules in assessing threats (e.g., see Beach, 1990; Lee, Mitchell, Wise, & Fireman, 1996; Lee, Mitchell, Holtom, McDaniel, & Hill, 1999). Such stop rules allow decision makers to evaluate fewer pieces of information when generating explanations for a decision environment. We tested Soldiers' use of stop rules when assessing the threat risk in images representing operational environments. We presented each image as a neutral environment and then added features (cues) to it over time. To each image, we added highly threat-relevant cues (e.g., disturbed earth, characteristic of improvised explosive device (IED) emplacement) and relatively innocuous cues (e.g., a change in the height of a wall). In some scenarios, the threat-relevant cue appeared early in the scenario, whereas in others it appeared later. Moreover, some scenarios represented familiar decision environments (e.g., detecting threats in an operational environment) and others represented unfamiliar decision environments (e.g., diagnosing medical conditions). Finally, we manipulated the amount of time Soldiers had to complete a fixed number of scenarios. To determine whether Soldiers employed cue-activated stop rules when assessing threat risk, we measured the number and type of cues Soldiers evaluated before identifying changes in their threat assessments.

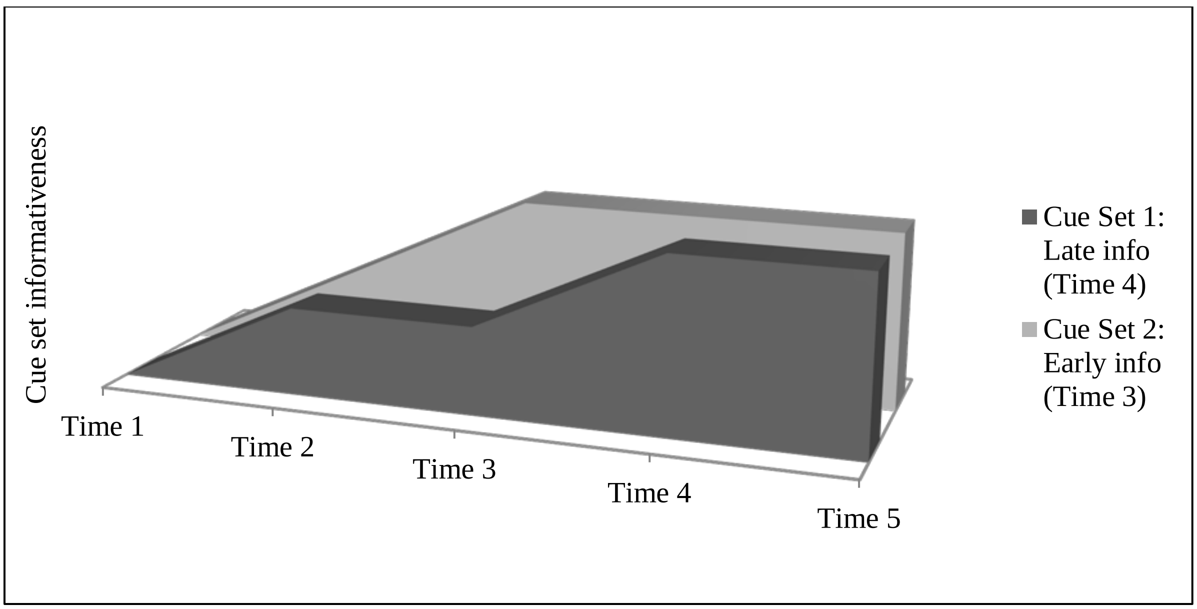


Figure 1. Cue-set informativeness over time.

Figure 1 depicts how we presented cues of differing informativeness in each scenario. The dark grey shaded area represents the informativeness of Cue Set 1, in which highly relevant information appears later in the sequence (Time 4). The light grey shaded area represents the informativeness of Cue Set 2, in which highly relevant information appears earlier in the sequence (Time 3). The best-informed and most efficient assessments will occur at Time 3 for Cue Set 2 and at Time 4 for Cue Set 1. If we consider that optimal timing involves changing an assessment in response to the most informative set of cues (see Figure 1), then a change in assessment that precedes peak informativeness may be suboptimal and potentially lead to lower quality decisions. Similarly, a change in assessment that occurs well after peak informativeness may involve assessing more information, but will also cost time and lead to inefficient decisions. We can consider decision efficiency at any time to equal the ratio of the combined informativeness of the set of cues evaluated up to and including that time over the elapsed time since cue set informativeness peaked. For example, in Figure 1, decision efficiency for each cue set at Time 5 may be calculated thusly:

<u>Cue Set 1</u>	<u>Cue Set 2</u>
Cue set value = 2	Cue set value = 2
Time at peak value (Time 5 – Time 4) = 1	Time at peak value (Time 5 – Time 3) = 2
Informativeness/time = 2/1 = 2	Informativeness/time = 2/2 = 1

Hence, decisions made at Time 5 will be more efficient for Cue Set 1 (efficiency = 2) than for Cue Set 2 (efficiency = 1)[1]. Poorly informed (early) or inefficient (late) decisions in operational environments can risk safety and cost lives. Thus, it is critical to understand how Soldiers balance the need to make accurate, informed decisions with the need to make efficient decisions.

In the first experiment of Leins et al. (In Preparation), we tested individual Soldiers and found a strong effect of time pressure on the timing of Soldiers' assessments. When under time pressure, Soldiers stopped scenarios and reported changes to their assessments sooner (i.e., they evaluated fewer pieces of information) than when they were under no time pressure. Consequently, time pressure influenced whether Soldiers changed their assessments at optimal or suboptimal times in each scenario. We speculated that when under no time pressure, Soldiers may have used a weighted-additive strategy to generate assessments. This strategy involves evaluating each successive cue until no more cues are available and then choosing an assessment (see Payne, Bettman, & Johnson, 1993). However, when under time pressure, and realizing they could not afford to use such a time-consuming strategy, Soldiers may have shifted to using a satisficing heuristic. Satisficing involves evaluating cues sequentially and stopping after finding a cue that surpasses a threshold of relevant informativeness (e.g., a criterion that helps categorize cues as high vs. low threat risk; see Gigerenzer & Goldstein, 1996; Simon, 1957). Given sequences of cues of varying informativeness, decision makers can lower their threshold of informativeness, allow a wider range of cue values to trigger an assessment, and choose a candidate assessment after evaluating less information than would be required by a higher threshold of informativeness. In our experiment, time pressure may have lowered Soldiers' thresholds of informativeness and allowed them to evaluate just enough cues to generate an effective (but not necessarily the best) assessment.

In the second experiment of Leins et al. (In Preparation), we explored whether Soldiers working collectively to assess threats would perform similarly to Soldiers working

---

[1] As explained later, we coded participant's changes in change assessments to indicate whether such responses were "early," "on-time," or "late." An "on-time" assessment change could result in a zero value for instance. Since we knew when the cue was introduced (early or late) and what image the participant was viewing at the time they changed their assessment, we could calculate the quality of the timing of their assessment.

independently. We tested a different sample of Soldiers, participating in groups of 3-4. These groups viewed the same stimuli as seen in Experiment 1, provided initial assessments for each scenario independently, but also discussed their threat assessments as groups. We did not observe a similar effect of time pressure on these groups. Rather, they reliably delayed changes to their assessments longer than individuals across all conditions. Hence, they were generally less efficient than individual Soldiers, but this may have protected them against making suboptimal assessments based on too little information. Of interest, however, we found a strong effect of task familiarity by which groups of Soldiers reliably stopped familiar scenarios earlier when the most relevant cue appeared early, but they evaluated more information in unfamiliar scenarios regardless of when the most relevant cue appeared. These groups appeared better able than individual Soldiers to leverage recognition-based heuristics. Perhaps groups were able to use recognition-based heuristics more effectively because those groups had broader pools of experience from which to draw: Given the same experience levels, four Soldiers' schemas together likely cover a larger and more diverse set of events than do the schemas of just one Soldier. This explanation is consistent with the notion that the ability to leverage recognition-based heuristics increases with greater experience and expertise in a domain (see Klein, 1993, 1997; Shanteau, 1992a, 1992b). What then might be the effect of shared domain-relevant experience on collective decision-making? We were interested in exploring whether the decision-making and heuristic usage of groups of Soldiers with shared deployment experience, and therefore shared domain knowledge, would differ from that of groups of Soldiers without similar shared experience. Thus, we designed, and describe here, an experiment in which groups of Soldiers with different levels of deployment experience engaged in collective decision-making.

Expertise in Decision-Making

Research has examined expert decision-making in various domains, ranging from military to healthcare to sports (Zsambok & Klein, 2014) and often found that it appears to be quicker and less effortful than novice decision-making. For example, Raab and Johnson (2007) found that expert handball players who generated predictions about a player's next move generated the same number of options as did novices, but arrived at those options more quickly. The recognition-primed decision (RPD) model suggests that these faster, less effortful judgments stem from experts' ability to recognize cues and engage in pattern matching (Klein, 1993, 1997). Through experience in a domain, experts develop deeper knowledge and a larger repertoire of relevant experiences they can retrieve from long-term memory. Drawing on these experiences, experts can quickly recognize potentially relevant information and generate explanations for a situation. This often looks like intuitive rather than effortful decision-making. For instance, expert firefighters have reported they spontaneously act rather than deliberately chose (Klein, Calderwood, and Clinton-Cirocco, 1986). This ability is particularly useful in time-pressured situations in which decision makers must find alternative or atypical solutions. Novices in these situations have less knowledge and fewer experiences to draw from, often resulting in slower and more deliberate hypothesis generation. However, as we observed in Leins et al. (In Preparation), working in groups can also slow decision-making and may lessen any advantage conferred by individual expertise. To explore this in the current experiment, we tested groups with different levels of deployment experience.

Similar to individual expertise development as a result of experience, groups can also develop a collective expertise. Groups may develop both task domain knowledge and group-specific knowledge. Similar to task domain knowledge, group-specific knowledge can lead to more efficient decision-making. One path to making group decision-making more efficient is to engage a transactive memory system (Wegner, 1987). Transactive memory refers to the division of cognitive labor among group members for the purpose of enhancing group memory capacity and performance. If two team members with similar memory capacity share the responsibility for storing and recalling the details of experiences, they increase their likelihood for jointly recalling those experiences more completely. Over time, group members develop implicit roles for encoding and recalling particular categories of details. These roles can develop naturally, for example, when one Soldier excels at encoding and recalling person details, while another Soldier may excel at encoding location details. Over time, the Soldier who best recalls person details may implicitly rely on the other Soldier to recall location details (and vice versa), thus removing the need to encode and/or recall both categories of detail. As a squad gains operational experience, members may naturally assume various roles aligning with their individual cognitive capabilities. During subsequent group decision-making, the group may divide the cognitive labor according to those implicit roles. In the following experiment, we sought to explore whether Soldiers with more experience working together would generate hypotheses more efficiently than Soldiers with less experience working together, and to see if potential efficiencies transferred across domains or if they were domain specific.

Design

We used a 2 x 2 x 2 x 2 mixed design with experience (relevant deployment experience vs. no relevant deployment experience) as a between-groups factor and cue-order (high-value first vs low-value first), familiarity (familiar vs. unfamiliar decision environment), and time pressure (low vs. high) as within-subjects factors. Each participant worked at a laptop computer while engaging in hypothesis generation tasks across experimental conditions. We manipulated the cue presentation order, the familiarity of the decision environment/task, and the time pressure associated with each decision task. To determine heuristic influence, we measured the number of images viewed before Soldiers reported a change in their hypothesis.

Independent Measures

We manipulated three variables across scenarios: order of cue values, familiarity of decision environment, and time pressure. We crossed both cue order and familiarity with time pressure. We nested cue order within familiarity.

Order of cues introduced. All scenarios presented incoming information (cues). Red arrows flashed onscreen identifying the location of new cue. Each incoming cue possessed either a high or low value of informativeness. A cue's value corresponded to how strongly it associated with a potential scenario status. A high-value cue in the threat detection context correlated strongly with a potential high threat risk according to subject matter experts (SME). A high-value cue in the medical diagnosis context correlated strongly with a high risk of a particular disease according to SMEs. In half of the scenarios, high-value cues appeared in position two in the cue sequence, whereas in the other half of the scenarios, high-value cues appeared in position three. Thus, the two orders of added cues were:

- Low-value cue early (Cue Set 1): Initial image → low-value cue added (position one) → low-value cue added (position two) → high-value cue added (position three) → low-value cue added (position four).

- High-value cue early (Cue Set 2): Initial image → low-value cue added (position one) → high-value cue added (position two) → low-value cue added (position three) → low-value cue added (position four).

We fully crossed cue presentation order with time pressure and familiarity.

Familiarity of decision environment. Half of the scenarios involved a familiar decision context, whereas the other half involved an unfamiliar decision context. Participants received general guidance on how to respond in these contexts. Familiar scenarios involved assessing the threat risk in an operational setting such as proceeding down urban and rural roads or paths or clearing buildings. They were told to consider five levels of threat risk when assessing each scenario: 1 = low/minimal risk, 2 = some risk, 3 = moderate risk, 4 = severe risk, 5 = very severe risk. Unfamiliar contexts included decisions associated with diagnosing medical conditions in individuals in a hospital triage environment. Participants considered the same 5-point risk scale when assessing the risk to each patient. The total set of experimental scenarios comprised five familiar context scenarios (including one practice scenario) and five unfamiliar context scenarios (including one practice scenario).

Time pressure. Each group completed four scenarios with low time pressure and four scenarios with high time pressure. Each set of high time pressure scenarios included instructions indicating that participants had only two minutes to complete 10 scenarios. In actuality, they had to complete only four scenarios in this set; participants were debriefed about this minor deception at the conclusion of the experiment. A digital numeric timer accompanied high time pressure scenarios. This timer counted down as time elapsed, so participants could monitor their status. Onscreen instructions notified participants prior to time pressure trials that (a) the timer paused while they typed responses and (b) they could stop a scenario and pause the timer at any time by pulling the trigger of their joystick.[2] Hence, participants could conserve time by providing assessments early in any time-pressured trial. The purpose of the high time pressure was to determine whether Soldiers would make decisions based on suboptimal cue values (i.e., low-value cues) when they may reasonably anticipate having access only to those cues. We crossed time pressure with context familiarity and cue order.

Dependent Measures

To determine whether heuristics influence hypothesis generation, we recorded and time stamped the following for both individual participants and groups:

[2] Pulling the joystick trigger would pause the scenario, until participants restarted the scenario. Though there was concern that allowing participants to pause a scenario could result in an overly long delay in the decision process, we tended to see the impact we expected. That is, the somewhat arbitrary time pressure manipulation still interacted with experience. Particularly, we were interested in whether a simple time pressure manipulation might impact the response to the onset of cues (of different values) as well as with experience; the effect is discussed in the Results section.

- initial threat ratings (before cues were added to a scenario)
- supporting reasoning for the initial threat rating
- confidence rating for the initial assessment
- a second threat rating (after cues were added to a scenario)
- supporting reasoning for the second threat rating
- confidence rating for the second assessment.

We coded secondary assessments for their timing, that is, whether participants stopped scenarios before, coinciding with, or after the appearance of the high-value cue. We also coded video recordings of group discussions for group dynamics, including group members' proportion of contribution to discussions.

Experimental Hypotheses

We predicted that groups of Soldiers with relevant deployment experience would generate hypotheses more efficiently than groups of Soldiers without relevant deployment experience. Specifically, experienced groups would provide a greater proportion of their reassessments coinciding with the appearance high-value cues (i.e., register more "on-time" reassessments) than would inexperienced groups (Hypothesis 1). In addition, we predicted that deployment experience would interact with cue order such that experienced groups would indicate reassessments sooner in response to early high-value cues and later in response to late high-value cues, but the timing of inexperienced groups' reassessments would not vary as a function of cue order (Hypothesis 2). We also predicted experience would interact with task familiarity such that experienced groups would be more likely than inexperienced groups to register on-time reassessments in familiar but not unfamiliar scenarios (Hypothesis 3). Finally, we predicted that experience would interact with time pressure such that experienced groups would be unaffected by time pressure, but inexperienced groups would accelerate their decision-making and deliver reassessments too soon when under time pressure versus when under no time pressure (Hypothesis 4).

Method

Participants

We tested 115 Soldiers in 30 groups of 3-4 (five groups of 3 members and 25 groups of 4 members). Sample 1 comprised 51 Soldiers in groups of 3-4. Sample 2 comprised the remaining 64 Soldiers in groups of 4. Overall, individual group members indicated a mean age of 22.68 years, a mean of 2.77 years in service, and a mean of 1 year in their current rank (see Table 1 for additional demographic data). The groups in Sample 1 represented groups of inexperienced Soldiers: groups of Soldiers with little or no relevant deployment experience and little or no time working together in the same unit. Sample 2 was composed of experienced Soldiers: groups of Soldiers with relevant deployment experience and time spent working together in the same unit. These two samples differed regarding the proportion of each group that had relevant deployment experience. In Sample 1, 16% of Soldiers had relevant deployment experience; these Soldiers presented randomly across groups so that, on average, only one member (27%) of each group had relevant deployment experience. By contrast, three of four members (75%) of each group in Sample 2 had relevant deployment experience, $t(28) = 5.22$, p <

.001. Soldiers participated in pre-formed groups. We assigned groups randomly to experimental conditions, counterbalancing the order of scenario presentation.

Table 1
Participant Demographics

		Sample 1		Sample 2	
		N	%	N	%
Current Rank	E-1:	5	10	2	3
	E-2:	11	22	5	8
	E-3:	8	16	19	30
	E-4:	17	33	35	54
	E-5:	9	18	3	5
	E-6:	1	2	0	0
Number of reported training courses aiding threat detection ability	0:	17	33	13	20
	1:	21	41	38	60
	2+:	13	26	13	20
Deployed	Yes:	8	16	53	83
	No:	43	84	11	17
Of participants who deployed:		n	%	n	%
Number of deployments	1:	6	75	42	79
	2:	1	13	8	15
	3+:	1	13	3	6
Number of times “outside the wire”	Never:	1	13	14	26
	< 1/month:	1	13	15	28
	1/month:	0	0	7	13
	> 1/month:	1	13	11	21
	1/week:	1	13	2	4
	> 1/week:	2	25	3	6
	Every day:	2	25	1	2
Ever deployed with current unit	Yes:	5	10	48	75
	No:	46	90	16	25

Note. The “deployed” variable includes only combat deployments (e.g., we did not code deployment to South Korea as a combat deployment). Current Military Occupation Specialties (MOS) reported by Sample 1 participants (n participants in parentheses): 11B(6), 11C(2), 12B(4), 12N(1), 13B(6), 13D(6), 13F(1), 13R(4), 25U(1), 35M(1), 42A(1), 56M(1), 88M(2), 91A(2), 91B(4), 91F(2), 91J(1), 91M(1), 91S(2), 92A(1), 92Y(1). Current MOS reported by Sample 2 participants: 11B (64).

Materials

Participants interacted with laptop computers running the Psychology Experiment Building Language (PEBL; Mueller & Piper, 2014) application. One laptop computer ran an application that presented all experimental stimuli and recorded all group responses in the form of joystick pulls, touchpad button presses, and keyed text. Up to four additional laptop computers ran an application that presented prompts for individual data, including codes to associate entered data with their corresponding experimental stimuli. For example, when participants entered an assessment for the third image in the fourth scenario, they entered the code 4:3 (4 = scenario number, 3 = image number). Participants also entered their individual threat ratings, reasoning, and confidence ratings using these laptops. Participants used Logitech Extreme 3D Pro joysticks to stop scenarios. These joysticks were paired with the laptop presenting experimental stimuli. Visual stimuli included scenarios presenting decision tasks and environments. Visual stimuli were projected onto a white projection screen. To capture group discussion, we recorded experimental sessions using a Panasonic digital camcorder, model HC-V270K.

Scenarios. Participants interacted with 10 scenarios. Five scenarios presented threat detection (familiar) tasks; the other five presented medical diagnosis (unfamiliar) tasks. One familiar scenario and one unfamiliar scenario were used as practice scenarios. Each scenario included a short description accompanied by a static image (see Appendix A for examples). Each description presented the scenario context and decision requirements (e.g., "Your squad is working through a village, classifying routes. Your squad leader has asked you to assess the threat level of this part of the route"). The description remained onscreen for 12 seconds, which was adequate time to read all the text. Following the description, participants were instructed to assess the scenario by providing a threat rating, a brief description justifying their rating, and a confidence rating in their assessment. The image remained onscreen throughout reporting. After participants entered their assessment data, a new cue appeared every six seconds until four new cues were added or until the participants stopped the trial to indicate a change to their assessment. We chose a six-second presentation time to allow participants adequate time to view new details, but also to advance each trial at a practical pace. Red indicator arrows accompanied all newly added cues to draw participants' attention to this incoming information. The arrows disappeared after one second and the new cue remained onscreen. The cues added over time varied in informativeness.

We consulted SMEs to determine cue informativeness values. We assigned low informativeness values to cues that military SMEs identified as low priority threats in the threat detection context and that medical SMEs identified as symptoms unspecific to a particular syndrome or illness in the medical diagnosis context. By contrast, we assigned high informativeness values to cues identified as high priority threats or identified as symptoms highly specific to a particular syndrome or illness. The first cue added to each scenario image was always a low-value cue. The second and third cues varied randomly in informativeness (high vs. low), with only one high-value cue added to each scenario and counterbalanced in its position (second vs. third) across trials. Thus, each scenario contained one high-value cue presented in either the second or third order position. We added a final low-value cue in the fourth order position in all scenarios. Table 2 below provides the basic trial by condition breakdown. After a group completed a scenario and indicated readiness to proceed (via button press), PEBL loaded the next scenario. This process continued until the group completed all 10 scenarios (including two practice scenarios). Researchers predetermined the scenario and condition order for a fixed sample of participants. The PEBL ran a separate script for each order according to a subject identification number keyed into the program at the start of a session. Participants were assigned randomly to subject identification numbers.

Table 2
Trial by Condition Breakdown

	T1	T2	T3	T4	T5	T6	T7	T8
Familiar/Unfamiliar	F	U	F	U	F	U	F	U
Time Pressure (High/Low)	H	H	H	H	L	L	L	L
High Value Cue Position	2	2	3	3	2	2	3	3

Demographic questionnaire and decision-making scales. After completing all 10 hypothesis generation scenarios, participants completed a demographic questionnaire and two decision-making scales (see Appendix B). The demographic questionnaire included questions about relevant military experience. The first decision-making scale participants completed was the Decision-Making Style (DMS) Scale (Scott & Bruce, 1995). This scale measures five types of decision-making style: rational, intuitive, dependent, avoidant, and spontaneous. According to Scott and Bruce (1995), rational decision makers search for, and logically evaluate, alternative hypotheses. Intuitive decision makers rely on hunches and feelings. Dependent decision makers search for advice and direction from others. Avoidant decision makers attempt to abstain from making decisions. Spontaneous decision makers possess a sense of immediacy and a desire to expedite the decision process. The second scale participants completed was a shortened Need for Cognitive Closure (NFCC) Scale (Roets & Van Hiel, 2011; for the original scale, see Webster & Kruglanski, 1994). This scale measured participants' dispositional desire to obtain answers on a given topic. Individuals high in NFCC prefer order, structure, and predictability versus disorder. They also possess a sense of urgency to reach swift decisions. We were interested in exploring whether any of these decision-making styles correlated with how much information participants would evaluate before generating new hypotheses.

Procedure

After consenting to participate, participants entered demographic information and then received general instructions on how to interact with the PEBL application. Each group member entered individual data into a laptop computer assigned to him or her. One group member was assigned to enter group data into a separate laptop computer. Group members were told that they would initially provide individual assessments of a scenario, then discuss those initial assessments as a group, reach a consensus, and enter a group assessment (threat rating, reasoning, and confidence rating). They would then monitor the scenario as it changed over time. They were told that they should pull their joystick trigger to stop a trial when they saw information that changed their initial assessment. They would enter their new threat rating, reasoning for the changed assessment, and report their confidence in the change.

After entering these data, they would discuss their reassessment with their group. The group would then decide if they would enter a new consensus assessment or opt to see more information. If they entered a new assessment, they would move on to the next scenario after entering those data. If they chose to see more information, the group member responsible for entering group data typed a special character ("@") into the group data laptop and the scenario would resume. After these instructions, participants completed a set of practice trials to become familiar with the procedure. Practice trials asked for hypotheses about one familiar and one unfamiliar context. Participants were then introduced to the timer used in high time pressure trials. Participants then received general task instructions according to their experimental condition (e.g., participants were told that they had either two minutes or no time limit to complete the next 10 scenarios).[3] Specific scenario instructions paired with relevant images followed the general instructions. For each scenario, participants entered a threat rating (1-5), their reasoning for the rating, and their confidence in their assessment (0-100). They discussed their assessment as a group, entered consensus data, and then watched for changes in the scenario. They pulled their joystick trigger to stop a trial when their assessment changed. If they did not pull the trigger, the trial concluded six seconds after the fourth cue appeared. At this time, participants received a prompt to report their threat rating, reasoning, and confidence. After entering these data, they once again discussed their assessments as a group, entered group consensus data, and advanced to the next scenario. After completing 10 scenarios (including practice), participants completed the DMS Scale and the NFCC Scale. Finally, researchers debriefed participants regarding the nature of the study, including the use of the minor deception to enhance the salience of the timer in time-pressured scenarios (in actuality, groups had two minutes to complete four trials), and answered any questions asked by participants.

Scoring

The PEBL application output all data into Microsoft Excel files. It populated each file with a subject identification (ID), data corresponding to each trial completed (i.e., the trial condition, the initial threat rating, hypothesis, confidence rating, the second threat rating, hypothesis, and confidence rating, and the serial order position of the image onscreen when the trial concluded), and all responses to demographic and scale questions.

We calculated the change in threat rating and confidence over time for each scenario by subtracting the value of the initial ratings from the value of the second ratings. Thus, a positive change in threat level rating reflected an increase in perceived threat. We also calculated the variance of the difference between individuals' initial ratings and their groups' consensus initial ratings.

Groups' reassessments were scored for the quality of their timing. We scored reassessments as early if they occurred prior to the appearance of the high-value cue, on time if they occurred in conjunction with the appearance of the high-value cue, and late if they occurred after the high-value cue had been supplemented by an additional low value cue. We also scored timing quality dichotomously, as either on-time (optimal) or not on-time (suboptimal).

To score for group member contribution, we reviewed each group's video and coded each group members' substantive statements and relevant non-verbal behavior. Substantive statements included suggestions regarding threat level (e.g., "I scored it a 4"). This also included scoring statements of agreement or disagreement and supporting arguments. Relevant non-verbal behavior included nodding to indicate agreement, head-shaking to indicate disagreement,

[3] Though participants, particularly in the low time pressure trials first group, might be tipped off to the time pressure manipulation, we still saw time pressure interact with experience for instance. Thus, even though our procedure for introducing the manipulation could have been stronger, we still achieved some effects from it. This minor deception was explained to all participants during the debrief.

and various hand and finger gestures to indicate agreement or numerical statements.[4] Statements not scored as contribution included irrelevant sidebars, meta-discussion about grammar, syntax, or spelling, and repetitions for the purpose of dictating items already discussed and agreed upon. We used these scores to calculate proportions of contribution for each participant in each scenario and used these proportion scores to calculate the contribution variance for each group. As group members contributed equitably to discussions, the group contribution variance decreased.

We scored data for all eight experimental scenarios. We averaged threat ratings, confidence ratings, and timing scores within a given condition. The following analyses included 30 groups, including one group that completed only two scenarios because of time constraints.[5]

In addition to examining the primary effects of cue order, task familiarity, time pressure and deployment experience, we were also interested in how the contribution of different group members might affect the decision process and the outcomes. Reimer and Hoffrage (2003) identified two group decision-making strategies relevant to this experiment: majority wins (consensus based on a group vote) and truth wins (consensus as a deferment to one member with knowledge of the solution). In familiar (threat detection) scenarios, we may expect to see a majority-wins strategy when groups comprise Soldiers of the same rank and experience. In these scenarios, we would expect to see equitable contribution across group members, multiple 'pro' and 'con' arguments supporting group members' proffered hypotheses, and some negotiation before settling on a final hypothesis. By contrast, when groups comprise Soldiers of differing ranks or experience, we may expect to see a truth-wins strategy, as lower ranking (or less experienced) Soldiers may defer to higher ranking (or more experienced) Soldiers. Under these conditions, we would expect to see fewer arguments and negotiations and less discussion prior to settling on a final hypothesis. In unfamiliar contexts, we might expect to see a greater proportion of majority-wins strategies, and more equitable contribution, because threat detection-specific experience and rank may be less critical in generating good hypotheses. However, given the salience of hierarchy in the Army, we may still see truth-wins strategies employed liberally even in unfamiliar contexts.

Results

Initial Threat Rating

Soldiers rated the initial threat risk of each scenario on a 5-point Likert-type scale (1= minimal or no risk, 2 = some risk, 3 = moderate risk, 4 = severe risk, and 5 = very severe risk). Overall, groups' mean initial threat rating for scenarios across conditions corresponded to a rating between "some risk" and "moderate risk" ($M = 2.61$, $SD = 0.68$). An independent samples t-test revealed no effect of experience on groups' initial threat ratings ($M_{inexperienced} = 2.38$, $SD = 0.76$ vs. $M_{experienced} = 2.81$, $SD = 0.55$), $t(28) = 1.77$, $p = .09$.

---

[4] In Leins et al. (In Preparation), we scored contribution via direct subjective scoring of video files. This was the most efficient method and was no less accurate than more resource intensive methods. Thus, we used the same method in this study.

[5] Because this group completed only two scenarios, some analyses excluded them and therefore used fewer degrees of freedom.

Task familiarity. A repeated-measures analysis of variance (ANOVA) revealed a main effect of task familiarity on initial threat ratings, whereby groups rated familiar scenarios as presenting a higher initial threat risk (M = 3.20, SD = 0.83) than the threat risk presented by unfamiliar scenarios (M = 1.91, SD = 0.54), $F(1, 27) = 138.15$, MS = 23.90, $p < .001$, *partial* η^2 = .837. Task familiarity and experience did not interact to influence initial threat rating, $F(1, 27) = 1.28$, MS = 0.22, $p = .27$, *partial* η^2 = .045.

Time pressure. A repeated-measures ANOVA revealed no effect of time pressure on initial threat ratings, $F(1, 27) = 1.43$, MS = 0.07, $p = .24$, *partial* η^2 = .05, and no interaction effect of time pressure and experience, $F(1, 27) = 2.68$, MS = 0.13, $p = .11$, *partial* η^2 = .09.

Individual-Group Differences

We calculated the extent to which individual group members' initial ratings differed from the initial consensus group rating for each scenario. Overall, groups tended to report consensus ratings that were similar to individual group members' ratings ($M_{difference}$ = -0.02, SD = 0.18). An independent samples t-test revealed no effect of experience on the mean difference between individual group members' initial threat ratings and their group's initial threat ratings, $t(28)$ = 0.28, $p = .783$.

Task familiarity. A repeated-measures ANOVA revealed a main effect of task familiarity on individual-group differences in initial ratings, whereby in familiar tasks, after deliberating, group consensus rating were slightly lower than individual group members' initial ratings ($M_{difference}$ = -0.14, SD = 0.27), whereas in unfamiliar tasks, group consensus ratings were slightly higher than individual group members' initial ratings ($M_{difference}$ = 0.11, SD = 0.20), $F(1, 22) = 20.69$, MS = 0.91, $p < .001$, *partial* η^2 = .434. Task familiarity and experience did not interact to influence individual-group differences in initial threat ratings, $F(1, 27) = 0.26$, MS = 0.01, $p = .62$, *partial* η^2 = .009.

Time pressure. A repeated-measures ANOVA revealed no effect of time pressure on individual-group differences in initial ratings, $F(1, 27) = 3.80$, MS = 0.12, $p = .06$, *partial* η^2 = .123, and no interaction of time pressure and experience, $F(1, 27) = 1.13$, MS = .04, $p = .30$, *partial* η^2 = .04.

Group Confidence

Overall, groups rated their confidence moderately high across all scenarios, M = 86.92, SD = 10.97. An independent samples t-test revealed no effect of experience on group confidence ratings, $t(28) = 1.26$, $p = .22$.

Task familiarity. A repeated-measures ANOVA revealed no effect of task familiarity on group confidence, $F(1, 27) = 0.45$, MS = 11.89, $p = .51$, *partial* η^2 = .016, and no interaction between experience and task familiarity, $F(1, 27) = 0.02$, MS = 0.52, $p = .89$, *partial* η^2 = .001.

Time pressure. A repeated-measures ANOVA revealed no main effect of time pressure on group confidence, $F(1, 27) = 0.04$, MS = 0.60, $p = .85$, *partial* η^2 = .001. However, time pressure interacted with experience to influence group confidence: Groups with experience reported higher confidence when under time pressure than when under no time pressure, whereas groups without experience reported lower confidence when under time pressure than when under no time pressure (see Figure 2), $F(1, 27) = 5.16$, $MS = 84.88$, $p = .03$, *partial* $\eta^2 = .161$.

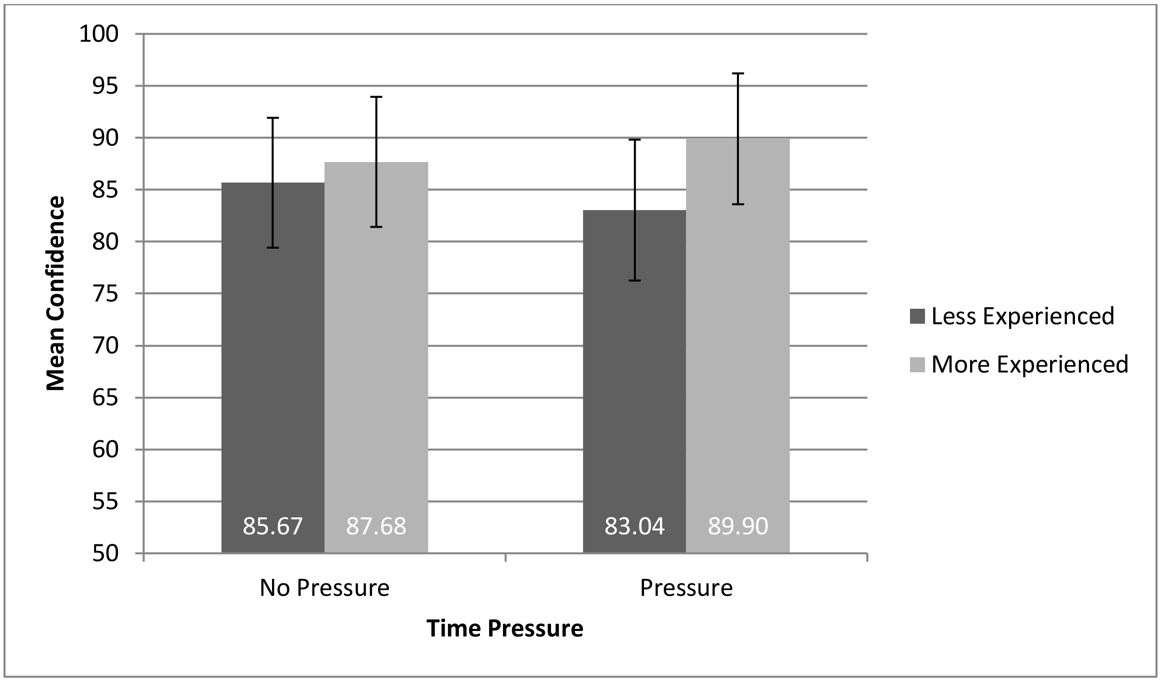


Figure 2. Interaction effect of time pressure and experience on mean confidence.

Change in Group Threat Rating over Time

Overall, groups tended to raise their threat ratings by a mean of 0.77 (SD = 0.30). An independent samples t-test revealed no effect of experience on changes in threat rating over time, $t(28) = 0.90$, $p = .38$.

Cue order. A repeated-measures ANOVA revealed no main effect of cue order, $F(1, 28) = 0.75$, $MS = 0.08$, $p = .39$, *partial* $\eta^2 = .026$, and no interaction effect of cue order and experience on changes in threat rating over time, $F(1, 28) = 1.62$, $MS = 0.17$, $p = .21$, *partial* $\eta^2 = .055$.

Task familiarity. A repeated-measures ANOVA revealed a main effect of task familiarity on changes in group threat rating over time, whereby groups' threat ratings increased by more in unfamiliar scenarios ($M = 0.99$, $SD = 0.40$) than in familiar scenarios ($M = 0.60$, $SD = 0.39$), $F(1, 27) = 12.44$, $MS = 2.22$, $p = .002$, *partial* $\eta^2 = .315$. However, task familiarity did not interact with experience to influence changes in threat rating over time, $F(2, 26) = 8.07$, $MS = .64$, $p = .002$, *partial* $\eta^2 = .383$.

Time pressure. A repeated-measures ANOVA revealed no effect of time pressure, $F(1, 27) = 2.59$, $MS = 0.18$, $p = .12$, *partial* $\eta^2 = .087$, or interaction between time pressure and experience, on changes in group threat rating over time, $F(1, 27) = 2.21$, $MS = 0.15$, $p = .15$, *partial* $\eta^2 = .076$.

Change in Group Confidence over Time

Overall, both experienced and inexperience groups tended to raise their confidence from Time 1 (initial rating) to Time 2 (reported change in assessment; $M_{difference}$ = 2.84, SD = 5.78). An independent samples t-test revealed no effect of experience on changes in group confidence over time, $t(28) = 0.74$, $p = .47$.

Cue order. A repeated-measures ANOVA revealed no effect of cue order, $F(1, 28) = 0.92$, $MS = 34.04$, $p = .35$, *partial* $\eta^2 = .032$, and no interaction of cue order and experience, $F(2, 28) = 0.46$, $MS = 16.84$, $p = .51$, *partial* $\eta^2 = .016$, on changes in group confidence over time.

Task familiarity. A repeated-measures ANOVA revealed no effect of task familiarity, $F(1, 27) = 0.14$, $MS = 4.62$, $p = .71$, *partial* $\eta^2 = .005$, or interaction between experience and task familiarity on changes in group confidence over time, $F(2, 27) = 0.33$, $MS = 10.82$, $p = .57$, *partial* $\eta^2 = .012$.

Time pressure. A repeated-measures ANOVA revealed no effect of time pressure, $F(1, 27) = 0.07$, $MS = 1.39$, $p = .79$, *partial* $\eta^2 = .003$, or interaction between time pressure and experience on changes in group confidence over time, $F(1, 27) = 1.23$, $MS = 23.50$, $p = .28$, *partial* $\eta^2 = .044$.

Quality of Timing of Changes in Assessments

In general, experienced groups stopped scenarios sooner than did inexperienced groups. Each scenario presented a series of up to nine images, depending on when the participant stopped the trial. Across all conditions, experienced groups tended to stop each scenario at image six, whereas inexperienced groups tended to stop each scenario at image seven, $t(28) = 2.64$, $p = .01$, $d = 0.96$.

We also calculated the proportion of scenarios that were stopped by groups too early (i.e., before the high-value cue appeared), on time (i.e., coinciding with the high-value cue), and late (i.e., coinciding with the introduction of cues after the high-value cue appeared). Across experience levels, groups were more likely to stop scenarios on time ($M = 0.37$, $SD = 0.20$) than early ($M = 0.21$, $SD = 0.22$), but no more likely to stop scenarios on time versus late, $F(2, 56) = 5.07$, $MS = .36$, $p = .01$, *partial* $\eta^2 = .143$. Deployment experience influenced the timing of changes in assessments. Groups of inexperienced Soldiers were more likely than experienced groups to stop scenarios late, $F(2, 26) = 8.07$, $MS = .64$, $p = .002$, *partial* $\eta^2 = .383$ (see Figure 3); thus, Hypothesis 1 was supported.

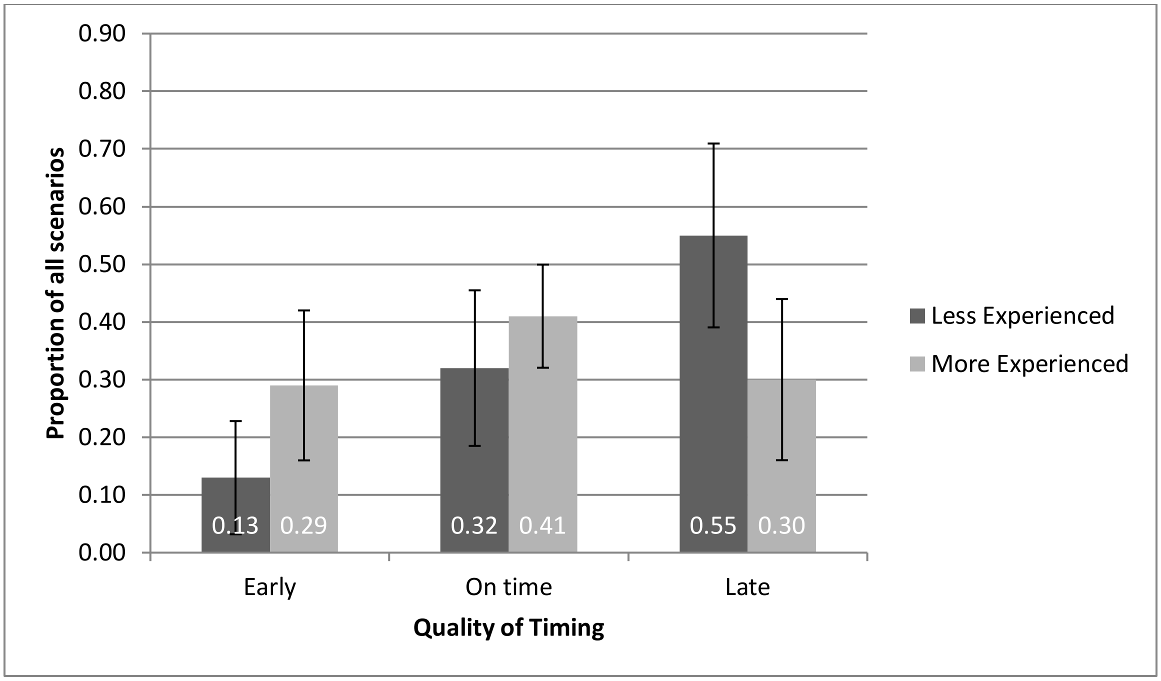


Figure 3. Influence of experience on timing quality.

Cue order. Examining the number of images viewed by groups, a paired-samples t-test revealed a main effect of order on when groups stopped scenarios, $t(29) = 2.36$, $p = .025$, $d = 0.43$. Groups stopped early-cue scenarios earlier ($M = 6.31$, $SD = 1.72$) than they stopped late-cue scenarios ($M = 6.89$, $SD = 1.31$). A repeated-measures ANOVA revealed no interaction between cue order and experience on the number of images viewed before providing a reassessment, $F(1, 28) = 0.60$, $MS = 0.56$, $p = .44$, *partial* $\eta^2 = .021$.

However, when examining only on-time assessment changes, a repeated-measures ANOVA revealed no effect of cue order on the quality of timing, $F(1, 28) = 0.05$, $MS = 0.003$, $p = .83$, *partial* $\eta^2 = .002$, and no interaction effect of cue order and experience on the quality of timing, $F(1, 28) = 0.69$, $MS = 0.05$, $p = .41$, *partial* $\eta^2 = .024$. Thus, although groups in general viewed fewer images in scenarios in which the high-value cue appeared early, they were still no more likely to make on-time reassessments. Moreover, experienced groups in general showed no more sensitivity to the timing of high-value cues than was demonstrated by inexperienced groups. Hypothesis 2 was not supported.

Task familiarity. Examining only on-time assessment changes, a repeated-measures ANOVA revealed no main effect of task familiarity on the proportion of on-time assessment changes, $F(1, 27) = 3.02$, $MS = 0.22$, $p = .09$, *partial* $\eta^2 = .101$. However, familiarity interacted with experience to influence the quality of timing, whereby experienced groups ($M = 0.53$, $SD = 0.27$) exhibited better timing than inexperienced groups ($M = 0.30$, $SD = 0.26$) during familiar scenarios, but not during unfamiliar scenarios, $F(1, 27) = 5.02$, $MS = 0.37$, $p = .03$, *partial* $\eta^2 = .157$ (see Figure 4); thus, Hypothesis 3 was supported.

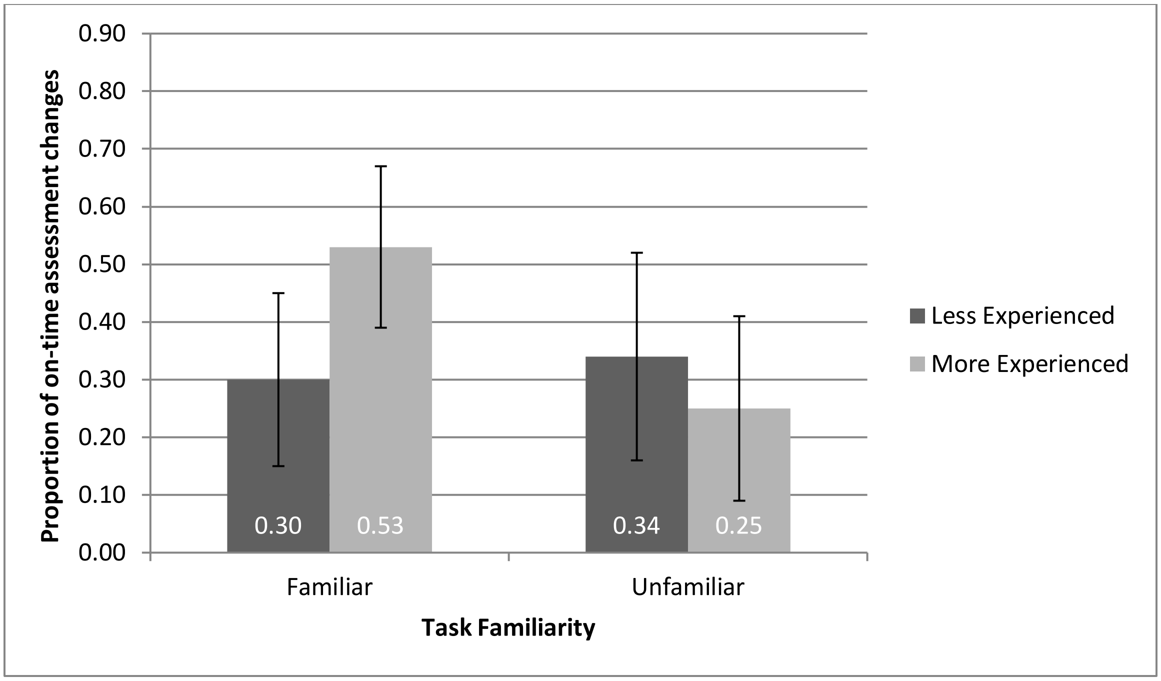


Figure 4. Interaction effect of experience and task familiarity on the proportion of on-time assessment changes. Note. Experience influenced timing within familiar scenarios, $t(28) = 2.4$, $p = .023$.

Time pressure. Overall, time pressure did not influence the timing of reassessments, $t(28) = 0.48$, $p = 0.64$. However, time pressure did interact with experience to influence the timing of reassessments. When under time pressure, groups differed in their distribution of timing across early, on-time, and late assessments. Whereas experienced groups were no more likely to reassess early, on time, or late when under time pressure, $F(2, 26) = 0.13$, $MS = 0.001$, $p = .99$, *partial* $\eta^2 = .001$, inexperienced groups were more likely to reassess scenarios late versus early or on time, $F(2, 26) = 6.40$, $MS = 0.86$, $p = .006$, *partial* $\eta^2 = .330$ (see Figure 5); thus, Hypothesis 4 was not supported. Moreover, time pressure did not interact with task familiarity to influence the quality of timing, $F(1, 27) = 0.05$, $MS = 0.003$, $p = .82$, *partial* $\eta^2 = .002$.

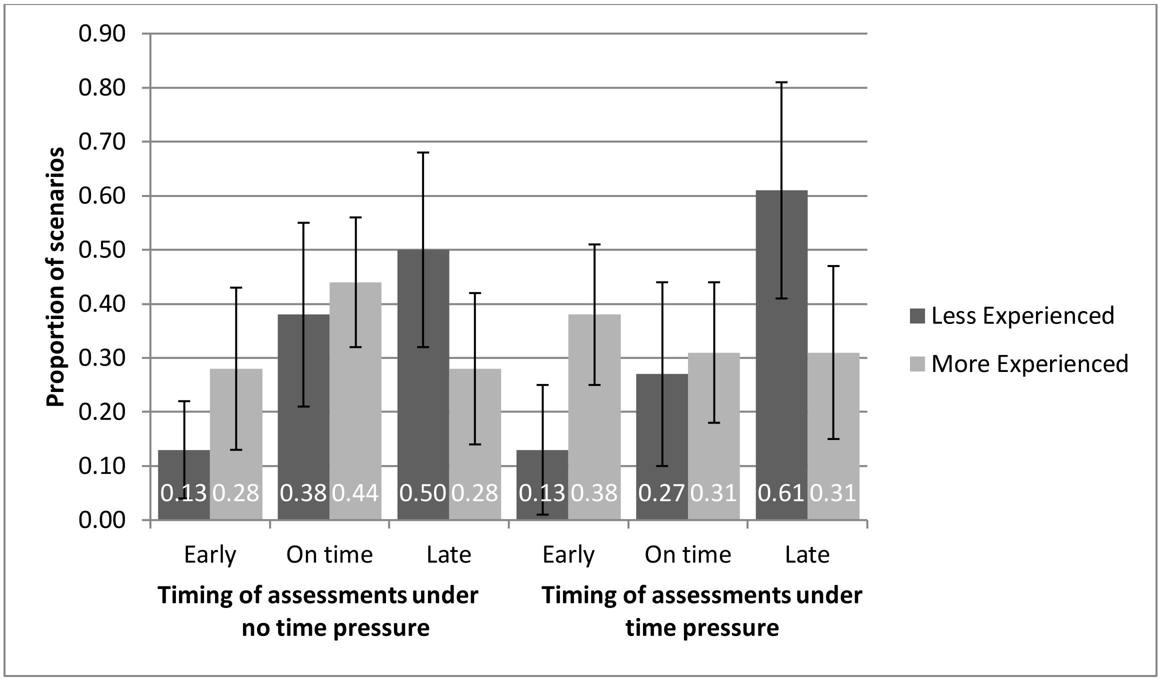


Figure 5. Interaction effect of pressure and experience on the proportion of early, on time, and late reassessments.

Three-way interactions. A related-samples Friedman Test using cue order, familiarity, and time pressure as independent variables and quality of timing (early, on time, late) as the dependent variable revealed no interaction effect on the quality of timing, χ^2 (7, 30) = 13.32, p = .07. This null effect held for both experienced and inexperienced groups, both $\chi^2 < 8.25$, both $p > .31$.

Examining only on-time assessment changes, a repeated-measures ANOVA using cue order, familiarity, and experience as independent variables revealed a three-way interaction effect on the quality of timing, $F(3, 81) = 3.87$, $MS = 0.38$, $p = .01$, *partial* $\eta^2 = .125$. When the scenario was familiar and the high-value cue appeared late, groups of experienced Soldiers were more likely to stop scenarios at optimal times versus when the scenario was unfamiliar and the high-value cue appeared late, $F(1, 27) = 8.16$, $MS = 0.31$, $p = .008$, *partial* $\eta^2 = .232$. There was no similar effect for experienced Soldiers when the high-value cue appeared early. By contrast, the timing of groups of inexperienced Soldiers did not differ as a function of the interaction of cue order and scenario familiarity, $F(1, 27) = 5.57$, $MS = 0.58$, $p = .026$, *partial* $\eta^2 = .171$ (see Figure 6).

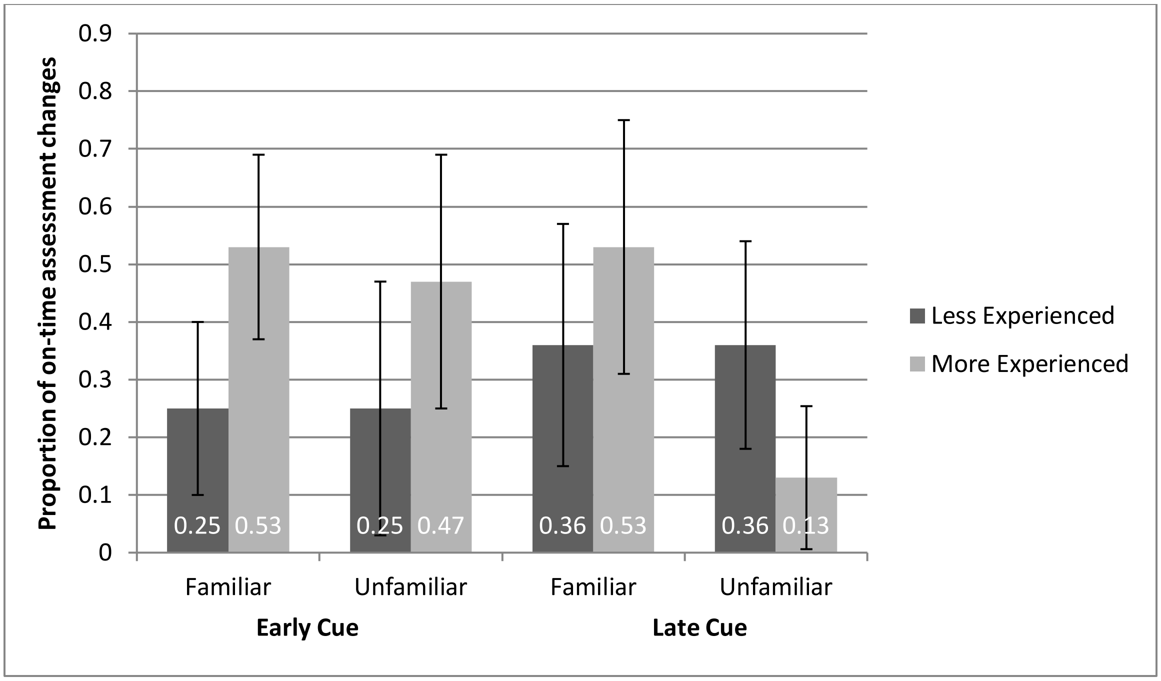


Figure 6. Cue order x familiarity x experience interaction effect on quality of timing.

Supplemental Analyses

In addition to analyzing the effects of cue order, task familiarity, time pressure, and deployment experience on threat assessments and the timing of those assessments, we explored whether groups differed regarding (a) how much group members contributed to consensus discussions (contribution variance) and (b) group decision-making traits as measured by group mean scores[6] on two decision-making disposition scales, the DMS scale and the NFCC scale. Two decision-making styles correlated with timing quality: As groups scored higher in either dependent or avoidant decision-making styles, they tended to stop scenarios later (dependent: r = -.29, p = .06; avoidant: r = -.32, p = .04). However, only the avoidant decision- making style approached statistical significance in a regression analysis predicting the timing of reassessments, B = -.201, t = 1.93, p = .06. No other decision-making style correlated with reassessment timing, nor did NFCC scores correlate with reassessment timing.

Contribution variance. An independent samples t-test revealed no difference in group contribution variance between experienced (*SD* = .07) and inexperienced groups (*SD* = .08), $t(28) = 0.62$, p = .54. Moreover, contribution variance did not correlate with the timing of assessment changes or the number of images viewed in any condition, all r values < .28, all p values > .13. Similarly, contribution variance did not correlate with any confidence measures, all r values < .22, all p values > .23.

[6] It might be argued that using the arithmetic group mean for the group scale analyses based on individual level response scales could lead to an unnecessary mixing of between and within-group variances. If the results of the scales indicated more effect or impact on results, we would recommend some sort of multi-level analytic approach that would allow a separation of individual and group level variances, such as hierarchical linear modeling.

Discussion

To summarize, we observed effects of all four independent variables: deployment experience, cue order, task familiarity, and time pressure. In support of our first hypothesis, groups of experienced Soldiers were generally more efficient at reassessing threats than were groups of inexperienced Soldiers. That is, experienced groups were more likely than inexperienced groups to stop a scenario at the time a highly threat-relevant cue appeared on screen, when the set of cues was most informative. By contrast, inexperienced groups were more likely to continue evaluating cues after the highly relevant cue appeared and well after the cue set's informativeness peaked (recall Figure 1). Moreover, in support of our third hypothesis, experienced groups exhibited better timing when scenarios presented a familiar task versus an unfamiliar task, whereas inexperienced groups were equally inefficient across task familiarity. This suggests that experienced groups were more likely than inexperienced groups to engage cue-activated stop rules that were better calibrated to the familiar decision tasks. These decision tasks included a mix of cues that subject matter experts (SME) differentiated as more or less relevant to threat detection. When experienced Soldiers perceived highly relevant cues in a familiar domain, those cues passed a threshold of informativeness, the Soldiers stopped searching for additional information, and they indicated a reassessment.

By contrast, inexperienced Soldiers often failed to differentiate the relevance of those same cues. Thus, high-value cues did not exceed the threshold at which inexperienced Soldiers would stop searching for information, and consequently, these Soldiers continued searching for and evaluating cues. They performed this way whether the scenario depicted a familiar task or an unfamiliar task. These results align with the notion that domain experience allows decision makers to leverage recognition-based heuristics when assessing domain-specific situations (Klein, 1993, 1997; Shanteau, 1992a, 1992b). Of interest, experienced groups and inexperienced groups did not differ in their threat ratings, either initially or after reassessing scenarios. It appears that experienced groups were simply more efficient at reassessing threat risk in dynamic scenarios representing operational environments.

In Leins et al. (In Preparation), we speculated that individual Soldiers who delayed their assessments beyond peak cue-set informativeness were using weighted-additive strategies to assess threat risk. That is, they evaluated cues sequentially until no more cues were available and then finalized their assessments (see Payne, Bettman, & Johnson, 1993). However, when they experienced time pressure, those same Soldiers seemed to abandon that strategy in favor of a more efficient strategy such as satisficing. Satisficing involves lowering the threshold of informativeness at which a cue will trigger a new situation assessment (Gigerenzer & Goldstein, 1996; Simon, 1957). Critically, however, we did not observe the same effect of time pressure on groups in that study. Rather, even under time pressure, groups continued to delay reassessments until after they evaluated all or most of the available cues.

In the current study, we observed mixed effects of time pressure. Overall, groups' proportions of on-time reassessments dropped from 45% of all reassessments under no time pressure to 35% of all reassessments under time pressure. However, upon further examination, inexperienced groups drove this effect in a surprising way: They delayed a majority of their reassessments when under time pressure. In fact, neither experienced nor inexperienced groups disproportionately accelerated their decision-making under time pressure. Thus, similar to Leins et al. (In Preparation), we found that groups did not make suboptimal early decisions and therefore likely did not shift to a satisficing strategy when under time pressure. Rather, they likely consistently applied recognition-based heuristics with uniform thresholds of informativeness across pressure and no-pressure conditions. This allowed experienced groups to continue to make a substantial number of on-time assessments, particularly on familiar tasks, but it did not similarly benefit inexperienced groups. Inexperienced groups, by nature, did not have the same experiences or rich memory traces to rely on when evaluating the informativeness of environmental cues and thus did not recognize high-value cues as highly threat relevant, regardless of task familiarity.

Further supporting the notion that experienced groups were better able to implement recognition-based heuristics in reassessing scenarios, we observed an interaction effect of experience, cue order, and task familiarity on the quality of timing of reassessments. Whereas inexperienced Soldiers appeared unable to recognize high-value cues at a rate much better than chance across task familiarity and cue order, experienced Soldiers were differentially affected by cue order and task familiarity. They were able to consistently register roughly half their reassessments on time in familiar scenarios, regardless of when the high-value cue appeared, but they were unable to maintain this rate across cue order in unfamiliar scenarios. Particularly, when the high-value cue appeared late in unfamiliar scenarios, experienced groups were less likely to register reassessments at optimal times. Essentially, experienced groups performed reasonably efficiently on familiar scenarios, regardless of when the high-value cue appeared, but were inconsistent in unfamiliar scenarios. By contrast, inexperienced groups performed inefficiently across all conditions. Again, this effect may be a result of groups of experienced Soldiers possessing multiple relevant schemas allowing for rapid matches of environmental cues to patterns of cues in memory. Groups of inexperienced Soldiers likely possessed fewer relevant schemas for assessing threat risk, and thus had delayed pattern matching or no pattern matching at all.

We observed several other effects unrelated to groups' deployment experience. For example, task familiarity influenced groups' initial threat ratings, the difference between individuals' and groups' initial threat ratings, and the change in threat ratings over time. Groups rated familiar scenarios as posing a greater threat risk than unfamiliar scenarios, but after deliberation, groups agreed that familiar scenarios were not as initially threatening as individuals alone had perceived them to be. In contrast, unfamiliar scenarios were in fact more threatening than individuals had originally perceived them to be. Thus, it appears that group deliberations served to reduce the polarity between individuals' threat ratings across scenario familiarity. Furthermore, although threat ratings generally increased over time, they increased disproportionately more for unfamiliar scenarios compared to familiar scenarios. We observed similar effects in Leins et al. (In Preparation), in which both individual Soldiers and groups of Soldiers reported greater initial threat risks for familiar scenarios than for unfamiliar scenarios, and in which both samples also reported disproportionately larger increases in threat levels for familiar versus unfamiliar scenarios over time.

It is unclear what mechanisms underlie these effects; however, it may be relevant to note that whether Soldiers evaluate an operational environment individually or collectively could influence their estimates of the environment's threat risk. This may be inconsequential unless the effect of collective decision-making is to reduce the estimate enough to induce Soldiers to reduce their vigilance and increase their actual risk. Given groups' assessments in this experiment, however, such a drastic adjustment seems unlikely. Even when assessing scenarios depicting relatively neutral operational environments, groups of Soldiers estimated the threat risk to be moderate. In fact, many individual Soldiers expressed the philosophy that operational environments will never pose anything less than a moderate risk. Nevertheless, future studies might benefit from further examination of how collective decision-making can moderate individual decision-making. We discuss one possible vein of this research below.

Regarding collective decision-making, similar to Leins et al. (In Preparation), we examined individual contributions to group consensus discussions, and we similarly found no influence of within-group contribution variance on any outcome. Furthermore, we found no influence of deployment experience on within-group contribution variance. We may have reasonably expected contribution variance to be greater among the inexperienced sample, as this sample also comprised groups made up of Soldiers of varying ranks and MOS. We had observed in Leins et al. (In Preparation) that when groups contained Soldiers of varying ranks, those with higher ranks tended to dominate discussions. We failed to replicate that effect in the current study. However, we also failed to observe any effect of task familiarity on the distribution of discussion across group members. If groups were to demonstrate the use of something like a transactive memory system (TMS) when deliberating, we might expect to see different levels of contribution across different tasks or domains. This was not apparent in the current study (however, see Appendix C for a figure representing the relative contribution of members of experienced groups across task familiarity). Perhaps the manipulations, scoring, or power to detect effects in this study were simply insensitive to the differences in communication that would arise with the use of a TMS. Specifically, groups of Soldiers who were not practiced in diagnosing medical conditions may not have developed a TMS relevant to that domain.

Ren and Argote (2011) suggest that critical factors impacting effective use of a TMS include group training, group member technical competence, depth of domain knowledge, and an understanding of who possesses what knowledge. Without prior exposure to the unfamiliar decision tasks in this study, groups of Soldiers simply may not have had the opportunity to develop those critical factors and therefore could not take advantage of a TMS when completing those tasks. However, they may have been able to leverage a TMS when completing familiar tasks. Ren and Argote (2011) also suggest that successfully using a TMS results in increased efficiency, an outcome we observed among experienced groups when they completed familiar decision tasks. Thus, a TMS may have subtly influenced the efficiency with which groups reassessed familiar scenarios. Unfortunately, our design does not allow us to disentangle some of the relevant antecedents of TMS usage. Experienced groups may have been more efficient because of an effective division of cognitive labor that would occur with use of a TMS or they may have been more efficient simply because their experience better allowed them to apply recognition-based heuristics. Future studies can help to better tease apart these explanations further explore Soldiers' use of transactive memory systems by carefully manipulating the knowledge possessed by each group member, as well as the knowledge required to solve experimental problems. To that end, different constellations of knowledge within a group should differentially confer advantages to that group across problems.

These findings can help to develop new research questions and to develop better measures of performance for home station training. If deployment experience allows Soldiers to use heuristics more efficiently in threat detection, and those gains in efficiency can be observed in decision-making, then Army leaders and trainers should be able to develop training exercises that provide learning opportunities similar to those provided by deployment experience. That could subsequently allow trainers to better evaluate Soldier and small-unit performance on decision-making in threat detection. For example, training developers can manipulate some of the same environmental variables we manipulated in this experiment. They can place environmental features of varying relevance in training lanes and allow evaluators to determine whether Soldiers prioritize those features appropriately and efficiently. Moreover, training developers could further explore how individual and collective decision-making influences critical training tasks and the steps that comprise those tasks. It may be at the subtask level that evaluators can measure Soldier and unit performance to better evaluate and remediate the decision-making that underlies critical tasks such as threat detection.

Conclusion

In this effort, we observed influences of deployment experience, cue order, decision task familiarity, and time pressure on the timing of groups' reassessments. In general, the findings indicate that when assessing threat risk collectively, groups of Soldiers likely rely on recognition-based heuristics for identifying threat-relevant features in an environment. Although we found previously that Soldiers working independently likely engaged different assessment strategies as a function of the context or decision space in which they operated, here we found no evidence in a change in strategy as a function of environmental conditions. Instead, we found a strong effect of domain familiarity on the efficiency with which groups assessed the threat risk in a decision environment. A familiar context and early access to valuable information allowed groups of Soldiers with domain experience to successfully engage heuristics and present assessments efficiently. In these cases, groups of Soldiers were able to limit the amount of information they needed to evaluate the decision space and make a good, informed assessment (e.g., see Gigerenzer & Brighton, 2009). The experiment presented here represents another step toward understanding how decision environments influence the way Soldiers use heuristics to generate hypotheses collectively. The findings imply that as Soldiers gain domain experience, they become more proficient at applying recognition based heuristics and more efficient at generating hypotheses. Future research and interventions should aim to understand more about how Soldiers can acquire domain-specific knowledge in training environments, and how leaders and trainers can promote the effective use of heuristics in detecting threats in operational environments.

References

Beach, L. R. (1990). *Image theory: Decision making in personal and organizational contexts.* New York: John Wiley & Sons.

Gigerenzer, G., & Brighton, H. (2009). Homo heuristicus: Why biased minds make better inferences. *Topics in Cognitive Sciences, 1,* 107-143.

Gigerenzer, G., & Goldstein, D. G. (1996). Reasoning the fast and frugal way: models of bounded rationality. *Psychological Review*, *103*(4), 650.

Klein, G. A. (1993). *A recognition-primed decision (RPD) model of rapid decision making*. Ablex Publishing Corporation.

Klein, G. (1997). The recognition-primed decision (RPD) model: Looking back, looking forward. In C. E. Zsambok & G. Klein (Eds.), *Naturalistic Decision Making* (pp. 285-292). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Klein, G. A., Calderwood, R., & Clinton-Cirocco, A. (1986). Rapid decision making on the fire ground. Proceedings of the Human Factors Society 30th Annual Meeting. 1, 576-580.

Lee, T. W., Mitchell, T. R., Holtom, B. C., McDaniel, L. S., & Hill, J. W. (1999). The unfolding model of voluntary turnover: A replication and extension. *Academy of Management Journal, 42,* 450-462.

Lee, T. W., Mitchell, T. R., Wise, L., & Fireman, S. (1996). An unfolding model of voluntary employee turnover. *Academy of Management Journal, 39*, 5-36.

Leins, D. A., Leonard, J., Zimmerman, L. A., Minchew, M. L., & Vowels, C. L. (In Preparation). The influence of decision environment on heuristic usage in individual and collective hypothesis generation. (ARI Research Report). Ft. Belvoir, VA: U.S. Army Research Institute for the Behavioral and Social Sciences.

Mueller, S. T., & Piper, B. J. (2014). The Psychology Experiment Building Language (PEBL) and PEBL test battery. *Journal of Neuroscience Methods*, *222*, 250-259.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Raab, M., & Johnson, J. G. (2007). Expertise-based differences in search and option-generation strategies. Journal of Experimental Psychology: Applied, 3, 158-170.

Reimer, T., & Hoffrage, U. (2003). Information aggregation in groups: The approach of simple group heuristics (SIGH). In R. Alterman, & D. Kirsch (Eds.): Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society (pp. 982-987). Boston: Cognitive Science Society.

Ren, Y., & Argote, L. (2011). Transactive Memory Systems 1985–2010: An integrative framework of key dimensions, antecedents, and consequences. *The Academy of Management Annals, 5(1)*, 189-229.

Roets, A., & Van Hiel, A. (2011). Item selection and validation of a brief, 15-item version of the need for cognitive closure scale. *Personality and Individual Differences, 50*, 90-94.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*(5), 818-831.

Shanteau, J. (1992a). Competence in experts: The role of task characteristics. *Organizational Behavior and Human Decision Processes, 53*, 252-266.

Shanteau, J. (1992b). How much information does an expert use? Is it relevant? *Acta Psychologica, 81*, 75-86.

Simon, H. A. (1957). *Models of man; social and rational*. Oxford, England: Wiley.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Wegner, D. M., (1987). Transactive memory: A contemporary analysis of the group mind. In Mullen, B., Goethals, G. R., (Eds.): Theories of Group Behavior (pp. 185-208). New York, NY: Springer.

Zsambok, C. E., & Klein, G. (2014). *Naturalistic decision making*. Psychology Press.

Appendix A

Scenario Examples

The following images are screen captures of stimuli presented in the PEBL application.

Image 1. General Instructions

You are about to view several images of situations that you will assess. You will be given specific instructions with each image. These instructions will indicate the type of assessment you need to make and how quickly you must make it. You will have 12 seconds to read each instruction.

You will see an initial image and then write your initial assessment of that image, identify the reasons for your assessment, and rate your confidence in your assessment on a scale of 0-100. Then the image will change over time. Information will be added to the image that may or may not help you further assess the situation.

You will pull your JOYSTICK TRIGGER if and when your assessment changes. You will then type a brief description of your assessment, and your confidence. You will press ENTER after typing your response and receive instructions for assessing the next situation.

You will now practice with two situations. LEFT MOUSE CLICK to continue.

Image 2. Familiar context: Initial image & decision task.

5:1

Your squad is tasked with classifying a route through a valley. Your squad leader has asked you to report on the threat level of classifying this route.

Image 3. Familiar context: Initial threat level entry screen

Image 4. Familiar context: High-value cue presented in serial position two

Note. The initial image depicted the valley minus the portions of the dirt road, the vehicle, and the red arrow (see Image 3 on page A-3).

Image 5. Familiar context: Second hypothesis entry screen

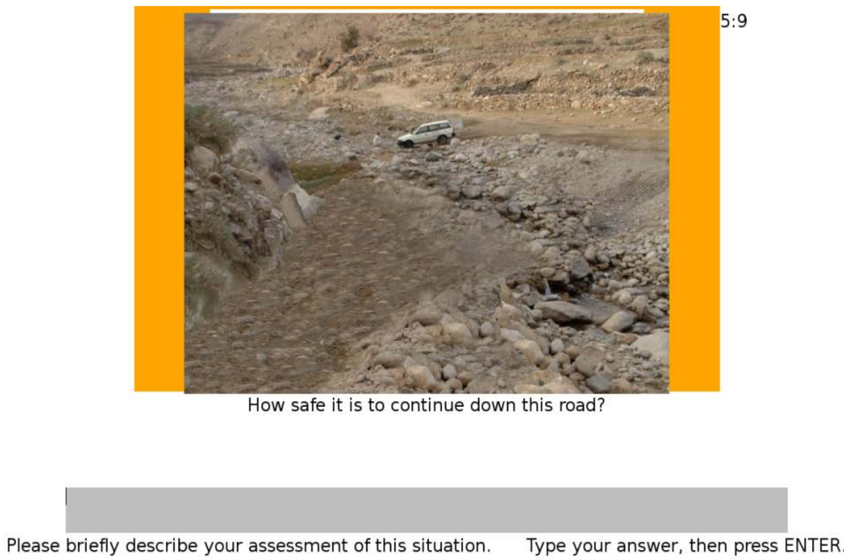


Image 6. Unfamiliar context: Initial image & decision task

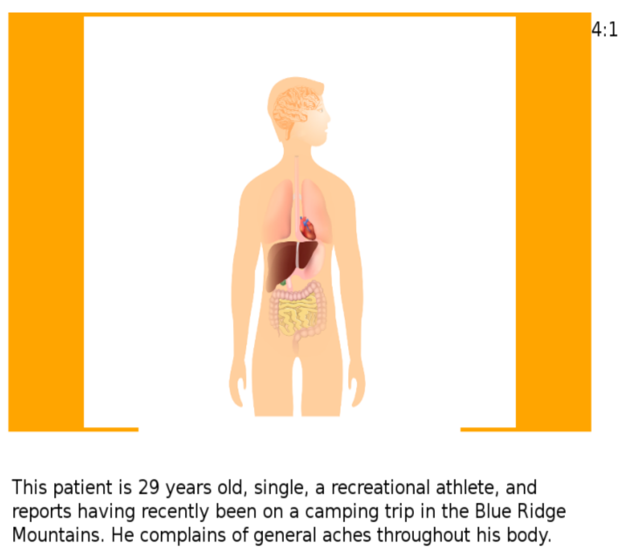


Image 7. Unfamiliar context: Initial threat level entry screen

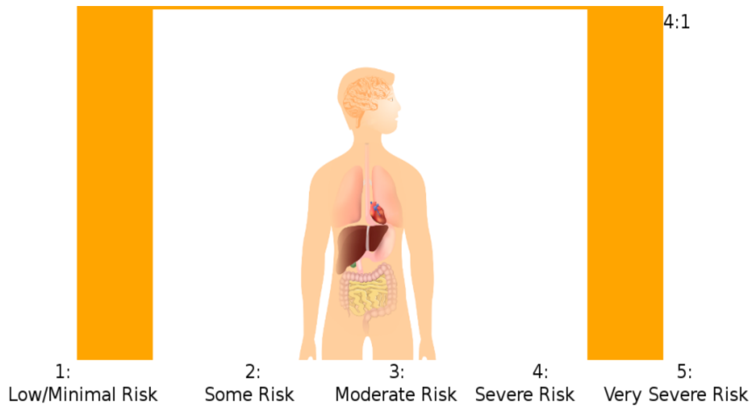
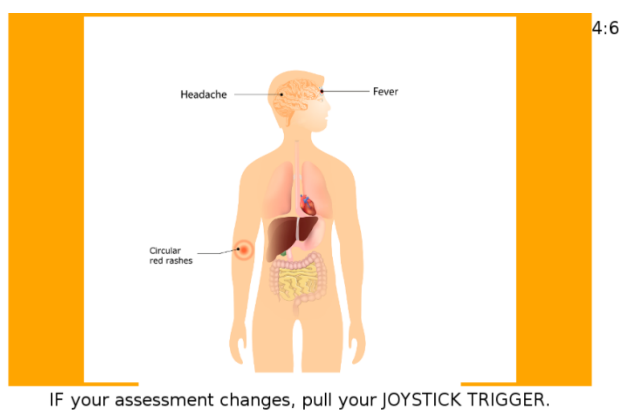
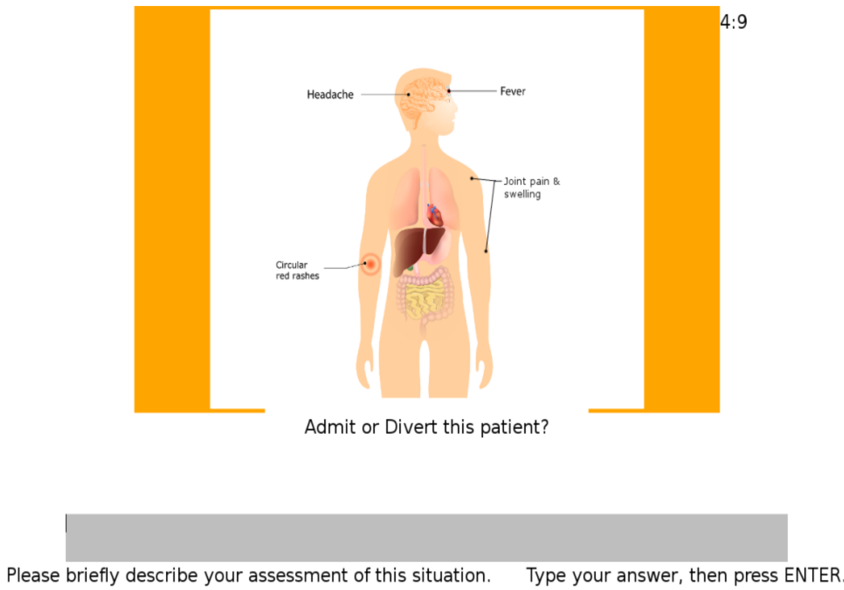


Image 8. Unfamiliar context: High-value cue presented in serial position three



Note. The high-value cue in this image is the collection of circular red rashes on the patient's arm.

Image 9. Unfamiliar context: Second hypothesis entry screen



Appendix B

Questionnaires and Scales

Demographic Questionnaire

1. Please enter your time in service, in years.
2. Please enter your current rank.
3. Please enter your time in current rank, in months.
4. Please enter your current MOS.
5. Please enter your age.
6. Have you ever deployed? Please answer with 'yes' or 'no.'
    a. If yes, how many times have you deployed?
7. Please enter the location of your most recent deployment (city or cities, and country).
8. Please enter your MOS at the time of your most recent deployment.
9. How often did you go 'outside the wire' on your most recent deployment? Please answer with: (0- never, 1- less than once a month, 2- once a month, 3- more than once a month, 4- Once a week, 5- More than once a week, 6- Every day).
10. Please describe any training you have received that improved your ability to detect threats and indicate the approximate date of the training month/year.
11. Please enter how long you have been with your current unit/squad. Enter your response in days, months, or years.
12. Have you been deployed with this current unit/squad? Yes/No

Decision-Making Style Scale

Scored on a 1-5 scale (1 = Strongly Disagree, 2 = Somewhat Disagree, 3 = Neither Agree nor Disagree, 4 = Somewhat Agree, 5 = Strongly Agree)

"Listed below are statements describing how individuals go about making important decisions. Please indicate how much you agree with each statement."

1. I double-check my information sources to be sure I have the right facts before making decisions.
2. I make decisions in a logical and systematic way.
3. My decision-making requires careful thought.
4. When making a decision, I consider various options in terms of a specific goal.
5. When I make decisions, I rely upon my instincts.
6. When making decisions, I tend to rely on my intuition.
7. I generally make decisions that feel right to me.
8. When I make a decision, it is more important for me to feel the decision is right than to have a rational reason for it.
9. When I make a decision, I trust my inner feelings and reactions.
10. I often need the assistance of other people when making important decisions.
11. I rarely make important decisions without consulting other people.
12. If I have the support of others, it is easier for me to make important decisions.
13. I use the advice of other people in making my important decisions.
14. I like to have someone to steer me in the right direction when I am faced with important decisions.
15. I avoid making important decisions until the pressure is on.
16. I postpone decision-making whenever possible.
17. I often procrastinate when it comes to making important decisions.
18. I generally make important decisions at the last minute.
19. I put off making many decisions because thinking about them makes me uneasy.
20. I generally make snap decisions.
21. I often make decisions on the spur of the moment.
22. I make quick decisions.
23. I often make impulsive decisions.
24. When making decisions, I do what seems natural at the moment.

Need for Cognitive Closure Scale

Scored on a 1-6 scale (1 = Strongly Disagree, 2 = Moderately Disagree, 3 = Slightly Disagree, 4 = Slightly Agree, 5 = Moderately Agree, 6 = Strongly Agree)

"Read each of the following statements and decide how much you agree with each according to your beliefs and experiences."

1. I don't like situations that are uncertain.
2. I dislike questions which could be answered in many different ways.
3. I find that a well ordered life with regular hours suits my temperament.
4. I feel uncomfortable when I don't understand the reason why an even occurred in my life.
5. I feel irritated when one person disagrees with what everyone else in a group believes.
6. I don't like to be with people who are capable of unexpected actions.
7. I don't like to into a situation without knowing what I can expect from it.
8. I dislike it when a person's statement could mean many different things.
9. I find that establishing a consistent routine enables me to enjoy my life more.
10. I enjoy having a clear and structured mode of life.
11. I do not usually consult many different options before forming my own view.
12. I dislike unpredictable situations.
13. When I have made a decision, I feel relieved.
14. When I am confronted with a problem, I'm dying to reach a new solution very quickly.
15. I would quickly become impatient and irritated if I would not find a solution to a problem immediately.

Modified Need for Cognitive Closure Scale from Roets and Van Hiel (2011).

Original Need for Cognitive Closure Scale developed and validated by Webster and Kruglanski (1994).

Appendix C

Participant Contribution across Familiar vs. Unfamiliar Scenarios

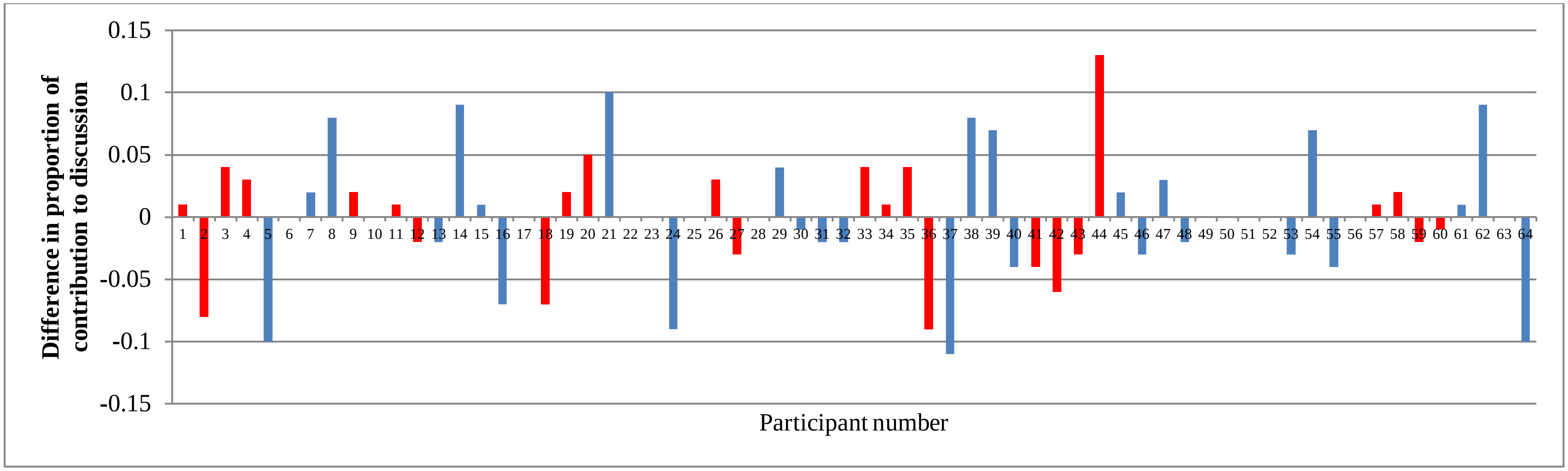


Figure C-1. Participant contribution data (Experienced Group, Sample 2). Note: Values represent the proportion of contribution to discussions of familiar scenarios minus the proportion of contribution to discussions of unfamiliar scenarios. Thus, positive values represent greater contribution to discussions of familiar scenarios; negative values represent greater contribution to discussions of unfamiliar scenarios; and, values of zero represent equal contribution across scenario familiarity. Red and blue colors alternate to signify different groups of four participants. Of interest, many groups contained one member who appears to have contributed disproportionately to familiar or unfamiliar scenario discussions. These are the types of distributions one might expect to see if groups relied disproportionately on different members for input regarding different domains – that is, if groups relied on transactive memory systems. However, the differences observed here (on the order of 5-10%) were too small to conclude that different group members dominated discussions across different domains.