19th ICCRTS, 2014 Paper Submission: Cover Sheet

Paper Number: 010

Title
Soldier Decision-Making for Allocation of Intelligence, Surveillance, and Reconnaissance Assets

Topics
Topic 4: Experimentation, Metrics, and Analysis (Primary)
Topic 3: Data, Information, and Knowledge (Alternate 1)
Topic 1: Concepts, Theory, and Policy (Alternate 2)

Authors
Jonathan Z. Bakdash
Human Research and Engineering Directorate, U.S. Army Research Laboratory

Diego Pizzocaro and Alun Preece
School of Computer Science and Informatics, Cardiff University

Point of Contact
Name: Jonathan Bakdash
Organization: Human Research and Engineering Directorate, U.S. Army Research Laboratory
Address:
ATTN: RDRL-HRS-E
459 Mulberry Point Road
Aberdeen Proving Ground, MD 21005-5425
E-mail: jonathan.z.bakdash@us.army.mil
Phone: (410) 278-9050

Distribution Statement

Soldier Decision-Making for Allocation of Intelligence, Surveillance, and Reconnaissance Assets

Jonathan Z. Bakdash
Human Research and Engineering Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, MD, USA
jonathan.z.bakdash@us.army.mil

Diego Pizzocaro and Alun Precee
School of Computer Science and Informatics
Cardiff University
Cardiff, UK
D.Pizzocaro@cs.cardiff.ac.uk
A.D.Preece@cs.cardiff.ac.uk

***Abstract*— Intelligence, Surveillance, and Reconnaissance (ISR) has been called the "… 'hub' of 21st century (military) operations." Military doctrine provides guidelines and protocols for ISR, but little is known about Soldier decision-making for the allocation of ISR platforms. To determine if technology may be useful for augmenting Soldier performance with ISR, we assessed the accuracy of decision-making using simulated allocation tasks. Soldiers made decisions by assigning ISR platform sensors to simplified target detection and identification tasks. The objective, or algorithmic accuracy of the decisions were based on the National Imagery Interpretability Reconnaissance Scale (NIIRS), which consists of normative ratings of imagery interpretability by intelligence analysts across varying sensor capabilities (i.e., pixels on the sensor). Algorithmic accuracy was derived from unclassified/open-source information on sensor capabilities based on NIIRS. Soldiers performed the same set of decision-making tasks twice. First, using their own knowledge and experience with ISR and, second, with complete information on sensor capabilities. Decision accuracy was slightly lower in the first set of assignments compared with the second. However, both were below algorithmic accuracy. Results indicate technology for decision aids with ISR allocation may enhance human decision-making.**

***Keywords*—** *Intelligence, Surveillance, and Reconnaissance; Decision-Making; Intelligence.*

I. INTRODUCTION

Intelligence, surveillance, and reconnaissance (ISR) has been called the "…'hub' of 21st Century (Military) Operations" [1]. ISR supports current and future military operations through the planning and operation of sensors and assets [2]. We focus on ISR allocation, which is the assignment of assets to target detection and identification tasks, for physical sensors on aerial platforms. Military doctrine on ISR provides extensive guidelines and protocols for the staff specific roles and responsibilities in ISR collection planning and the tasking of ISR resources [3]. However, little is known about actual Soldier decision-making for ISR allocation. One exception is research examining simulated ISR allocation for multiple assets, threats, and varying priority targets [4]. In contrast, we focus on decision-making for specific target detection and identification tasks.

How can technology help with ISR sensor allocation? To determine if technology is needed to enhance Soldier performance for ISR allocation, we investigated decision-making for sensor allocation for simulated target detection tasks.

An illustration of ISR allocation is described in the following vignette (adapted from [5]):

> A patrol notices a suspicious black car with license plate ABC123 moving south. A database query reveals that this vehicle is known to be associated with a high value target, John Smith. They lose sight of the vehicle. An intelligence analyst must decide which unmanned aerial vehicle (UAV) to allocate to find the car.

Most UAVs will likely have sufficient quality visual sensors to detect a black vehicle, but may not be able to distinguish between different types of cars, let alone identify the license plate. Thus, the ISR platform(s) capable of detecting the car will depend on whether or not it is necessary to read the license plate and an assortment of other factors.

ISR allocation has a complex problem space with interactions among social and natural systems, natural systems, and technical systems [6]. In the real-world, decisions for ISR allocation and the effectiveness of ISR may depend upon (list adapted from [4: pp. 1]):

1. *Social and natural systems*: Individual humans and groups (military, civilians, and insurgents), priorities such as force protection, Information Requirements, scheduled collection tasks, skill of the pilot or UAV platform operator, stress and fatigue, and time pressure.
2. *Natural systems*: Environmental characteristics: Current and future weather conditions, terrain, and time of day.
3. *Technical systems*: Sensor capabilities and platform capabilities, such as: speed, range, total flight time, and visual and acoustic detectability from the ground. These factors can also depend on natural systems. For example, flying into high wind reduces the speed, range, and flight time of an aerial platform.

Given the wide range of factors that can be involved in ISR allocation and the goal of determining if decision-making needs to be enhanced, we developed a simplified task to measure objective decision-making. Moreover, objective measurement is crucial to assessing actual human performance, especially in safety critical work domains, because subjective measures (e.g., observation, interviews, and preferences) can produce divergent data [7].

The ISR allocation task here had objectively correct or incorrect assignments. Specifically, the task involved deciding if sensors on different platforms were capable of performing target detection or identification tasks. Because the ISR allocation task was objective and there was no time pressure, the hypotheses were motivated by the theory and empirical findings of Actuarial Judgments (see Table 1, 4. Actuarial Judgments). This theory posits objective methods (i.e., statistical formula or algorithms) for decision-making generally have greater accuracy than subjective, human judgments. We hypothesized the following:

1. Decision-making will be more accurate with complete information on sensor capabilities
2. Decision-making accuracy will be below algorithmic accuracy, despite having complete information available

Results weakly supported the first hypothesis; a medium effect size for improvement in decisional accuracy was found with complete information. The second hypothesis was strongly supported, with a large effect size: Decision accuracy was below perfect algorithmic accuracy despite providing complete information.

A. Theories of Decision-Making

There is little research on empirical human decision-making for ISR allocation, but there are several major theories of human decision-making [8] and some common ground among theories [9], [10], [11]. Key differences between theories include the role of expertise and deviations, or lack thereof, from rationality. Because of these clear divisions there is no singular, unifying theory of human decision-making. Five major theories of human decision-making are described in Table 1.

TABLE 1. THEORIES OF DECISION-MAKING

Theory	Primary Discipline(s)	Description	References
1. Naturalistic Decision-Making (NDM)	Human Factors	Experts make decisions based on intuition and analysis. The Recognition-Primed Decision model is part of NDM. From experience, experts form patterns that can be used to quickly make decisions without having to evaluate all options.	Klein [12], [13]
		Primarily based on qualitative real-world data using observations and interviews. Limited quantitative lab data.	
2. Prospect Theory, also called Heuristics and Biases	(Behavioral) Economics and Psychology	Frequent systematic errors in human decision-making, interpreted as deviations from rationality due to systematic heuristics and biases in human decision-making. Two systems for decision-making, System 1 is slow and controlled and System 2 is fast and automatic. System 2 is heuristic based, which is consistent with NDM. Primarily based on quantitative lab data.	Kahneman and Tversky [8], [14]
3. Bounded Rationality and Fast and Frugal Heuristics	(Behavioral) Economics and Psychology	To make complex decisions, humans use heuristics: Simple search, satisficing/stopping (i.e., “good enough”), and other decision rules. These heuristics are adaptive with respect to the environment. Similarities to NDM and System 2 in Prospect Theory, with exceptions. For example, there are some situations where novices perform better than experts. Also, this theory suggests that some findings in Prospect Theory are, at least, partially attributable to the representation of information (percentages vs. natural frequencies such as 1 out of 100) rather than the actual decision-making process. Based on quantitative lab and quantitative	Simon [15]; Gigerenzer [9], [16]

Theory	Primary Discipline(s)	Description	References
		real-world data.	
4. Actuarial Judgments; also called Algoritmic or Statistical Judgements	Computer Science, Psychology, and Statistics	Actuarial or algorithmic/statistical decisions are generally more accurate than subjective human decisions. This is not so much a theory of human decision-making as a theory of fallibilities in human decision-making and the value of objective decision-making in many situations. Quantitative evidence from the lab and real-world: Disease diagnosis in health care, diagnosis and risk in clinical psychology, prediction of success in education, and investment performance.	Meehl and others [17], [18]
5. Game Theory	Computer Science, (Traditional) Economics, Mathematics, and Statistics	Generally assumes humans are rational to mathematically model human decision-making [19], with some exceptions [20]. Optimization of utility function(s) with respect to constraints. This theory is the same as Actuarial Judgments, except for the key assumption that human decision-making is rational, and therefore is accurately modeled by mathematical or statistical optimality. Weak support based on quantitative lab data and some support from real-world data, such as pricing and auctions.	Numerous researchers; for examples see [19], [21]

In the first three theories, decision-making may not be rational, and thus not mathematically optimal; hence, they are at odds with Game Theory. Actuarial Judgment and Game Theory are distinguishable by only one aspect: Actuarial Judgment is a theory of optimal objective decision-making, but does not claim to be an accurate model human judgment. Game Theory is generally used as a model to explain human decision-making under the assumption of rationality [19].

We based the hypotheses below on the theory of Actuarial Judgment for four main reasons. First, it has a clear implementation: using objective methods to enhance decision-making. This matches our goal of determining if technology, arriving at recommended decisions computationally, is needed to enhance human decision-making. Second, there is over six decades of empirical research supporting Actuarial Judgment with findings in a wide range of domains and this work has shown that even when the algorithm and the human have the same data, the algorithm is almost always more accurate [22]. Third, our simplified ISR allocation task had no time pressure nor did it have all of the complex information likely to be present in the real-world. Therefore, our task was not likely to be amenable to the pattern recognition of NDM or the heuristic accounts of Theories 2 and 3. Last, Game Theory has repeatedly been shown to be an inaccurate model of actual human decision-making, see [8].

B. Paper Structure

The reminder of the paper is structured as follows: Section II describes the ISR allocation task and statistical results and Section III has a discussion and conclusion, with recommendations using technology to enhance ISR decision-making.

II. ISR Allocation Task

In this section, we discuss the subject matter expert Soldiers, the study procedure and materials, and the study results.

A. Subject Matter Experts

Eleven U.S. Army Soldiers with operational ISR experience were recruited as subject matter experts (SMEs). One Soldier was excluded because he indicated on a survey questions that he did not have operational experience with ISR, only experience with ISR during training. SMEs consisted of nine males and one female. Soldiers had deployed experience with ISR ranging from management, collection, and analysis to direct experience with the ground effects of ISR. The rank, Military Occupational Specialty (MOS), and deployed experience of the SMEs are described in Table 2.

TABLE 2. MILITARY BACKGROUND OF SUBJECT MATTER EXPERTS

Rank[a]	Military Occupational Specialty[b]	Deployed Experience[c]
CPT	35D	BN Intelligence OIC

Rank[a]	Military Occupational Specialty[b]	Deployed Experience[c]
CPT	35D	Platoon Leader, BN Assistant Intelligence OIC, BN Intelligence OIC, Intel/Operations Combat Advisor, and WMD Coordination Intelligence Officer
CPT	35D	BDE Collection Manager
1LT	11A	Intelligence Advisor to Host Nation
1LT	11A	Intelligence Advisor to Host Nation
1LT	35D	BN Intelligence OIC and Intelligence Advisor to Host Nation
SSG	35F	DIV Intelligence Operations Analyst and BDE Collection Management
SSG	29E	BN Electronic Warfare SGT
SGT	35F	BN Intelligence OIC, Targeting NCO, and Current Operations Analyst
SGT	35F	BDE ISR Operations NCOIC

a. Rank descriptions: http://www.army.mil/symbols/armyranks.html

b. Military Occupational Specialty descriptions: www.apd.army.mil/Home/Links/PDFFiles/MOSBook.pdf

c. The descriptions of operational experience are generic to protect personally identifiable information. Acronyms for military echelons (unit sizes) are: DIV, BDE, BN, and CO, which respectively stands for Division, Brigade, Battalion, and Company. For a detailed description of military echelons, see http://en.wikipedia.org/wiki/Military_unit#Commands.2C_formations.2C_and_units OIC stands for Officer in Charge. NCOIC stands for Non-Commissioned Officer in Charge.

Note a., b., and c. in Table 2 were taken verbatim or with minor modifications from [4: Table 1, pp. 2]. Seven out of 10 Soldiers were trained intelligence analysts (35-series MOS), 2 were light infantry (11-series MOS), and 1 specialized in offensive electronic warfare (29-series MOS).

Table 3 has descriptive statistics on age, military service, and military deployments.

TABLE 3. DESCRIPTIVE STATISTICS OF SUBJECT MATTER EXPERTS

Variable	Mean	Standard Deviation
Age (years)	27.10	4.46
Military Service (years)	5.50	3.13
Deployments (number of times)	1.30	0.48

B. Procedure

SMEs were recruited using two methods:

1. Umbrella Week: This is a scheduled week in which units set aside times for the research and development community to interview Soldiers and administer surveys.
2. Asking other researchers and Soldiers for suggested contacts.

There was considerable difficulty finding qualified SMEs. The a priori projected sample size was N = 15–20 to meet or exceed 80% statistical power for a large effect size with paired sample t-test and default assumptions, calculated using *G*Power 3.1.7* [23]. However, we were only able to find 10Soldiers with operational ISR experience. In the sample, most Soldiers with relevant experience were intelligence analysts; however, we estimate, based on our recruitment and the expert opinions of Soldiers that ~1–1.5 per 100 Soldiers are intelligence analysts and further estimate 1 out of 10 intelligence analysts have ISR experience: This meant that only about 1–1.5 per 1,000 Soldiers met the study inclusion criteria. Repeated measures using multiple assignment decisions were used to increase statistical power.

Soldiers were told that participation was completely voluntary, that they could withdraw at any time and for any reason, and responses were non-attributional. SSMEs received no compensation for their participation. After completing the decision-making task, 6 out of 10 Soldiers also participated in interviews to assess Human Factors in ISR (see [6]). The first author administered paper or electronic questionnaires with the simulated ISR allocation tasks. Six Soldiers participated in person and four Soldiers received verbal instructions and then sent their responses over email.

C. Materials and Study Design

SMEs were told the purpose of the project was to look at decision-making for target detection using ISR. In addition, they were instructed to: (1) decide which sensor(s) on ISR platforms were good enough or better than needed to detect a target, (2) assume optimal conditions (ideal weather, time of day, and angle) (3) typical range for target detection, and (4) ignore platform speed. Study materials are available from: http://thedata.harvard.edu/dvn/dv/jbakdash/faces/study/StudyPage.xhtml?globalId=doi:10.7910/DVN/25583&studyListingIndex=0_598175c3d39df6b1cc38e2dc1de0

Objective sensor capabilities were derived from unclassified/open-source information based on the National Imagery Interpretability Rating Scale (NIIRS); see http://www.fas.org/irp/imint/niirs.htm. NIIRS is an empirically validated scale based on the accuracy of human analysts for assessing the normative quality of data from different physical sensors for a variety of target detection and identification tasks. In addition, NIIRS and sensor capabilities were also used to determine objective or algorithmic accuracy, this constituted perfect performance.

Soldiers performed the same set of decision-making tasks twice. First, Soldiers performed the task relying on their own knowledge and experience with ISR and, second, with the complete information on sensor capabilities and the criteria used to describe NIIRS. The questionnaire was structured as follows:

1. Demographic, military background, and military experience questions
2. **Set 1:** Sensor assignment decisions based on knowledge and experience for eight target detection tasks; confidence rating of overall decisions and strategy used to make decisions
3. **Set 2:** Sensor assignment decisions based on NIIRS for the same eight target detection tasks; overall confidence rating of decisions and strategy used to make decisions

In each set of assignments, SMEs completed 13 decisions on sensor assignments (for five assets) for the eight target detection tasks; a total of 104 decisions for each set and 208 decisions total. SMEs completed the questions at their own pace, taking 15 –45 minutes to finish the entire task.

The order of Set 1 and Set 2 was always fixed. Set order was not counter-balanced because providing NIIRS ratings could have biased decisions solely based on knowledge and experience. SMEs were permitted to look at their Set 1 decisions for their Set 2 responses, but were told not to change their answers to Set 1. Two examples of target detection tasks and the required sensor capabilities, shown in parentheses, are:

1. Known location, detect and identify the license plate on a vehicle (requires a Visual NIIRS rating 9; note no asset had a visual sensor capable of performing this task)
2. Moving car, jeep, or Humvee (requires a Visible NIIRS rating 4 or higher, Radar NIIRS 4 or higher, or IR NIIRS 5–6 or higher; note that all assets had sensors capable of performing this task)

Five different ISR platforms were available, platforms had visible, infrared (IR), and/or radar sensors. ISR platforms were selected based on the availability of unclassified/open-source information on sensors. The availability of sensor information determined the platforms; this is a limitation because of the similarities in NIIRS ratings. Table 4 shows the information provided to SMEs for Set 2 decisions, the NIIRS ratings of the five ISR platforms.

TABLE 4. UNCLASSIFIED/OPEN-SOURCE ISR PLATFORM NIIRS RATINGS

Platform Type	NIIRS Rating	Sensors
Predator A (MQ-1)	• Visible NIIRS rating 6 • IR NIIRS rating 6 • RADAR NIIRS rating 6	• EO/IR Camera • SAR
Reaper (MQ-9)	• Visible NIIRS rating 8 • IR NIIRS rating 8 • RADAR NIIRS rating 6	• EO/IR Camera • SAR
Raven	• Visible NIIRS rating 6 • IR NIIRS rating 6	• EO/IR Camera
Global Hawk	• Visible NIIRS rating 8 • IR NIIRS rating 8 • RADAR NIIRS rating 8	• EO/IR Camera • SAR
Shadow 200 (RQ-7)	• Visible NIIRS rating 7 • IR NIIRS rating 7	• EO/IR Camera

Note that the NIIRS were derived from actual values or estimates published in open-source and unclassified information, such as specification sheets, technical papers, and scientific papers; values were received via personal communication [24]. The NIIRS ratings are believed to be current as of January 2013.

SMEs were told verbally the information may not match classified capabilities or current sensors on platforms, but to still rely on the provided NIIRS ratings. Algorithmic accuracy is based on these NIIRS ratings.

D. Results

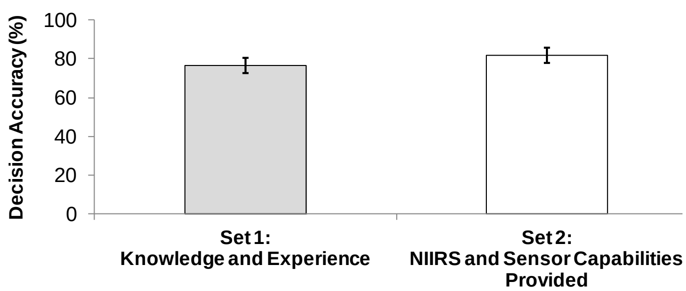
Data were analyzed using a paired sample t-test and a one-sample t-test. Accuracy for Set 1 and Set 2 was determined using mean value, across detection tasks and sensors, by SME. Accuracy was comprised of hits (correctly assigning a sensor capable of detecting the target) and correct rejections (correctly *not* assigning a sensor that was incapable of detecting the target). Individual SMEs made a total of 208 allocation decisions: 104 decisions for each set. However, the overall sample size was small: N = 10.

Because of the small sample, bootstrapping was used to calculate the statistical parameters for decision-making: t-test values, standard errors, and effect sizes and their confidence intervals. For small sample sizes, bootstrapping has better properties: (1) lower bias (absolute error in the estimator, i.e., the test statistics) and (2) greater efficiency (comparative effectiveness of the estimator for the given data relative to other estimators) for parameter estimation than conventional statistical methods that do not use resampling [25].

Bootstrapping is a data simulation method using random sampling without replacement for parameter estimation [26]. Analyses were performed using *R* [27] with bootstrapping implemented using the *boot* library [28]. One thousand bootstrap iterations were run for each t-test. The raw data and *R* code for reproducing the analyses are available from the above link for the study materials.

As hypothesized, a bootstrapped paired sample t-test showed that decision accuracy for ISR assignments was slightly lower for knowledge and experience (*Mean* = 76.50%, *SE* = 4.06) compared with full information on NIIRS (*Mean* = 81.60%, *SE* = 3.92), $t(17.98) = 1.85$, $p < 0.05$ (one-tailed), $d = 0.59$ (95% CI: 0.04 - 2.84 percentile bootstrap), see Figure 1.

FIGURE 1. DECISION ACCURACY USING KNOWLEDGE AND EXPERIENCE (SET 1) VS. COMPLETE INFORMATION (SET 2)

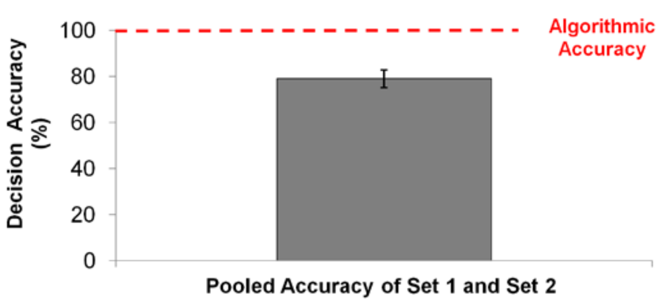


Error bars represent one bootstrapped standard error of the mean.

The medium effect size should be interpreted with caution because of the wide range of its confidence interval; the lower bound of the confidence interval nearly reaches zero. Nevertheless, the results suggest that complete information, albeit with high uncertainty, weakly improves the accuracy of decision-making.

A bootstrapped one-sample t-test (compared with 100%) indicated that pooled decision accuracy for ISR assignments (Set 1 and Set 2 combined) was lower (*Mean* = 79.05%, *SE* = 3.75) than algorithmic accuracy of 100%, $t(9) = 5.59$, $p < 0.001$ (one-tailed), $d = 1.77$ (95% CI: 1.42 - 4.23 percentile bootstrap), see Figure 2.

FIGURE 2. POOLED DECISION ACCURACY VS. ALGORITHMIC ACCURACY



Error bar is one bootstrapped standard error of the mean. The red dashed line indicates algorithmic accuracy (100%).

Again, due to the small sample, the range of confidence interval on the effect size is wide. However, the lower bound clearly exceeds a large effect size. One could argue that nearly 80% accuracy is reasonably good performance, but there was no time pressure, the task was simplified, and the information in Set 2 was sufficient for perfect performance.

Exploratory Results. Exploratory analysis was performed on the free response and subjective questionnaire data, see the Appendix for further details. Descriptive statistics, rather than inferential statistics, were used to examine this data because there were no a priori hypotheses. The exploratory results are summarized as follows:

1. *Allocation task:* Accuracy and errors varied between allocation tasks, suggesting differences in task difficulty.
2. *ISR Assets:* Accuracy was comparable across ISR assets.
3. *Free response questions:* In the first set of tasks, most Soldiers self-reported that they relied on their experience. In the second set, most Soldiers stated they relied upon the NIIRS ratings.
4. *Likert scale questions:* Overall, Soldiers indicated moderate experience with ISR platforms, weak experience with NIIRS, moderate confidence in their assignments for both sets and moderate use of assignment decisions made in Set 1 for Set 2 assignments; this was somewhat inconsistent with the free response data for decisional criteria, a reliance on just NIIRS was commonly reported. Last, more Soldiers reported that a system for ISR sensor assignments would often be helpful.

III. DISCUSSION

First, we discuss the possibility of combining actuarial judgments, as a form of partial automation, with human decision-making. Second, we cover human computer collaboration more generally. Third, we describe the Sensor Assignment to Missions (SAM) system [29], [30], which may be useful for enhancing human decision-making in ISR. Last, we explain limitations and possible future directions for the present work.

A. Actuarial Judgements, Automation, and Human Decision-Making

Research on actuarial judgments has shown repeatedly that the algorithmic method will outperform subjective human judgments the majority of the time [17]. These results cover a diverse range of decisions: diagnosis and treatment in health care, diagnosis and risk in psychology, education success, investment performance, and parolee recidivism [17], [18], [22]. However, this does not mean all human decision-making should be automated because there may be information that is obvious to the human but not incorporated in the algorithm, and novel situations that are out of the bounds of computation [18]. Automation raises clear safety concerns. Over and inappropriate automation has resulted in catastrophic accidents, including aircraft crashes and railroad accidents [31]. Decision aids can cause automation complacency and bias, where the humans may fail to properly monitor systems and/or the environment [32].

In safety critical domains, human supervisory control over technical systems is necessary to reduce the risk of accidents and loss of life [33]. Therefore, for ISR, we propose that algorithms provide transparent (i.e., rationale for system decisions) recommended decisions to Soldiers. This claim is bolstered by a finding that performance in simulated ISR tasking, for coverage and route planning, was enhanced by reliable, transparent automation under high task demands that involved multiple goals and constraints [4]. Similarly, computer assisted decision-making is superior to either humans alone or a computer alone for weather forecasting [34] and is often better for playing chess [35].

B. *Human Computer Collaboration*

Another approach is human computer collaboration (HCC), in which the human and one or more intelligent systems or agents work together with a common goal [36]. This approach is more interactive than computer assisted decision-making. A sizeable amount of work in this area has been conducted in relation to visual analytics, addressing analytic tasks the size and complexity of which make them intractable without close interplay of human and machine agents [37]. Recent work in the area of information fusion for ISR tasks has explored the use of controlled natural language for mission support, facilitating the interaction of human analysts with machine agents [5]. In terms of collaboration, in [38] the authors note that intelligence analysts are now well-versed in modern collaboration environments and social networking. The general notion that including social collaboration, and more broadly HCC, can improve the outcome of intelligence analysts is highlighted in [39]. There are both benefits and challenges in social collaboration and HCC challenges: "A richly collaborative environment, whether social, HCC, or both, could be a blessing, if computers can help sort, filter, and manage vast amounts of information, or a curse if volume of information is simply increased." [40, p. 12]

There are additional concerns with the implementation of HCC that are unique to safety critical domains, especially if even some degree of human supervisory control is ceded. For example, what if the human and computer disagree? What if the computer increases the likelihood of biases in human decision-making? Despite these concerns, there are compelling fictional examples of HCC for a collaborative and interactive interface [41] and computers can facilitate social collaborations.

C. *Sensor Assignment to Missions System*

One implementation of algorithmic judgments in ISR is SAM, a prototype artificial intelligence (AI) system [29], [30]. To transparently represent information, SAM builds on previous work [42] using an algorithmic assignments founded on the Military Missions and Means Framework (MMF) [43]. Information is formally represented using ontologies. *Missions* are comprised of *operations* that are in turn comprised of *tasks*. *Tasks* require *capabilities*, which are provided by *assets*. *Assets* include *platforms* and *systems*; *systems* – including *sensors* – are mounted on *platforms*. The relationship *allocatedTo* captures that an *asset* is assigned to resource a particular *task*. The interface for SAM on a mobile device is shown in Figure 3.

FIGURE 3. SAM IPAD INTERFACE

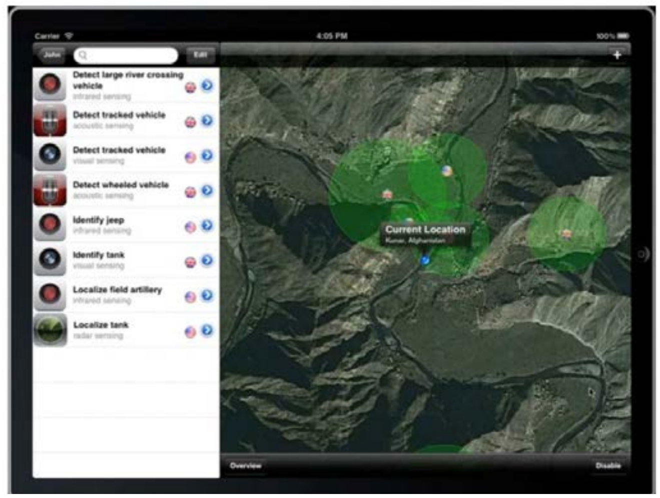


Image from [36: p. 9]

The ontology is implemented in the Web Ontology Language, OWL DL, and is shown in Figure 4.

FIGURE 4. MISSION AND MEANS FRAMEWORK FOR ISR ONTOLOGY

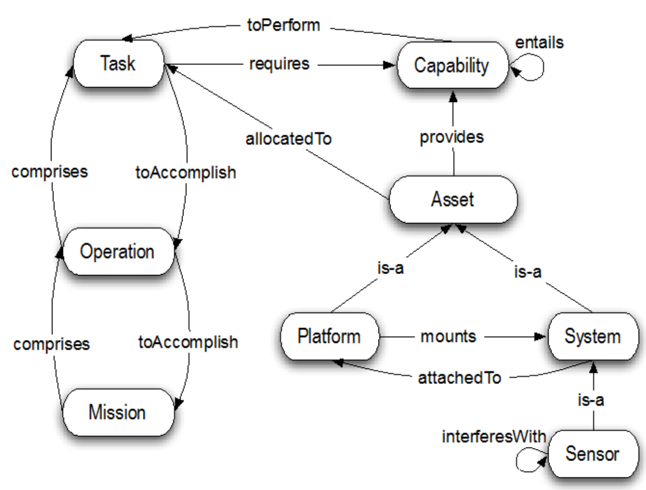


Image from [29: p. 4]

Sensor capabilities and detection tasks are characterized using NIIRS. Therefore, given an ISR task and a set of sensing assets in a particular area of interest, SAM provides the algorithmically optimal solution for allocating ISR resources. In addition, SAM for example, is capable of allocation based on the bearing and range of a platform to a task [45], in addition to matching NIIRS capabilities with task ISR requirements (via reasoning algorithms). Another potential application for SAM is training for ISR allocation based on NIIRS for sensors platforms and detection tasks.

An interactive conversational interface is being developed; this will allow non-programmers such as intelligence analysts to modify and update information [44]. With the conversational interface, Soldiers could refine and update the knowledge and SAM adding the sensor capabilities necessary to detect or identify new enemy tactics (e.g., putting an improvised explosive device [IED] on a donkey and sending it towards a checkpoint). Furthermore, Soldiers would have the capability to edit the optimal solutions found using algorithmically approaches to add that previously only reflected in human knowledge. The prototype conversational interface extends beyond computer assisted decision-making. Instead, human-computer collaboration is implemented through closed-loop feedback between the human and the intelligent system, see Figures 5 and 6.

FIGURE 5. CONCEPTUAL ILLUSTRATION OF CONVERSATIONAL INTERFACE FOR SAM

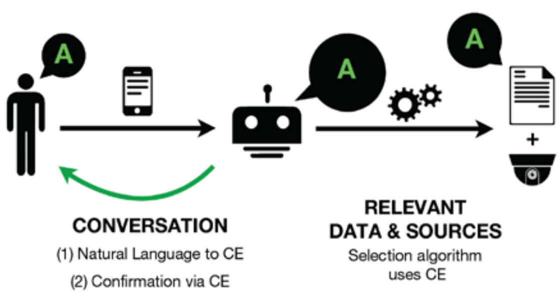


Image from [36: pp. 7]

FIGURE 6. CONCEPTUAL ILLUSTRATION OF CONVERSATIONAL INTERFACE FOR SAM

Image from [36: pp. 10]

D. Limitations

The work was unclassified, thus the open-source derived sensor capabilities may not have matched actual capabilities. This limitation is somewhat mitigated by providing Soldiers with NIIRS ratings for platforms in their second set of allocation decisions. Another limitation is that signals intelligence (SIGINT), which is highly classified, was not included in the ISR allocation task. Anecdotally, multiple Soldiers have stated SIGINT is often highly valuable because it leads to actionable intelligence more often than other types of intelligence. In addition, the task does not address the challenges or benefits of technical and human information fusion in the intelligence cycle, cross-cuing (using multiple sensor platforms to detect or identify targets), and allocation decisions for coordination among multiple ISR platforms and multiple collection tasks.

Last, the simplified but well-controlled research design of the task has weaknesses and strengths. The task did not incorporate multitude of factors that may be present in real-world ISR decisions, such as balancing multiple priorities, weather conditions, terrain, travel time, and skill of the pilot or platform operator (discussed in the Introduction in detail). A few Soldiers candidly remarked that the task was artificial, because of the many factors mentioned above. Although this statement is true, this controlled research design permits stronger inferences about the results than methods commonly used in real-world research: for example, observation which can be highly confounded [46] or verbal reports which can be subject to response bias [47]. Finally, the sample was not large enough to analyze individual differences in MOS, experience, or expertise.

The ISR task was designed for maximizing the accuracy of human decision-making and it only involved simple assignments for detection and identification. Nevertheless, sensor assignments for detection and identification are one dimension of ISR allocation. ISR coverage time and route (re)planning efficiency are other key aspects that have been previously investigated [4].

IV. CONCLUSION

The quantitative results in this paper provide supporting evidence for conclusions drawn in previous quantitative research on ISR coverage and planning [4] and qualitative work examining Human Factors issues in ISR [6]. The same recommendations we made previously, also apply here:

> "In unpredictable, dynamic work domains (such as ISR), we contend that enhancing human performance requires technical systems that are adaptive, interactive, integrated (as few unique systems as possible), and transparent (see [48], [49]). Decision aids may enhance Soldier decision-making for ISR allocation and resource management, but new technical capabilities need to also be flexible (e.g., ad-hoc and unofficial ISR requests) [6: p. 4]."

ISR is fundamental to military operations. We found weakly increased allocation accuracy when complete information on task relevant platform capabilities was provided. More importantly, even with complete information, decision accuracy was below algorithmic accuracy. This is quantitative evidence of a need for technology to enhance human decision-making with ISR. SAM has potential to be that technology, but ultimately further empirical research is needed to determine how to implement computer assisted decision-making in ISR.

The effectiveness of ISR depends on many factors. Some are uncontrollable factors (e.g., natural systems such as the weather and terrain), but decisions for ISR allocation are controllable. Ultimately, enhancing human-decisions for ISR using an implementation of computer assisted decision-making may increase the effectiveness of ISR and in turn improve the outcome of military operations.

Acknowledgments

We thank the Soldier SMEs for their time and thoughtful and insightful input and thank an anonymous Soldier for the donkey IED example. Also, we acknowledge Geeth de Mel for his knowledge of NIIRS and for providing unclassified/open-source ratings of sensor capabilities and are appreciative of recommendations made by Paul Sullivan regarding task instruction wording. We are grateful to Ellen Haas, Debbie Patton, Diane Ungvarsky, and several anonymous Soldiers for their expert advice and assistance in finding SMEs. Lastly, we thank Mark Gatlin for his assistance editing this paper.

This research was sponsored by the U.S. Army Research Laboratory and the U.K. Ministry of Defence and was accomplished under Agreement Number W911NF-06-3-0001. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Research Laboratory, the U.S. Government, the U.K. Ministry of Defense, or the U.K. Government. The U.S. and U.K. Governments are authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

References

[1] D. Deptula, “Air Force ISR in a Changing World: Changing Paradigms While Optimizing ‘Low Density’ to Meet ‘High Demand,’” presented at the American Institute of Aeronautics and Astronautics, 2010.

[2] “Joint Publication 1-02: Department of Defense Dictionary of Military and Associated Terms,” Jan. 2011.

[3] “FM 3-55: Information Collection,” Department of the Army, May 2013.

[4] E. Rovira, A. Cross, E. Leitch, and C. Bonaceto, “Displaying Contextual Information Reduces the Costs of Imperfect Decision Automation in Rapid Retasking of ISR Assets,” *Hum. Factors*, Jan. 2014.

[5] A. Preece, D. Pizzocaro, D. Braines, D. Mott, G. de Mel, and T. Pham, “Integrating hard and soft information sources for D2D using controlled natural language,” in *Information Fusion (FUSION), 2012 15th International Conference on*, 2012, pp. 1330–1337.

[6] J. Z. Bakdash, D. Pizzocaro, and A. Preece, “Human Factors in Intelligence, Surveillance, and Reconnaissance: Gaps for Soldiers and Technology Recommendations,” presented at the Military Communications Conference, 2013.

[7] A. D. Andre and C. D. Wickens, “When Users Want What’s not Best for Them,” *Ergon. Des.*, vol. 3, no. 4, pp. 10–14, 1995.

[8] D. Kahneman, *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux, 2011.

[9] G. Gigerenzer and R. Selten, *Bounded Rationality: The Adaptive Toolbox*. Boston: MIT Press, 2002.

[10] D. Kahneman and G. Klein, “Conditions for Intuitive Expertise: A Failure to Disagree.,” *Am. Psychol.*, vol. 64, no. 6, pp. 515–526, 2009.

[11] K. V. Katsikopoulos, T. Pachur, E. Machery, and A. Wallin, “From Meehl to Fast and Frugal Heuristics (and Back): New Insights into How to Bridge the Clinical--Actuarial Divide,” *Theory Psychol.*, vol. 18, no. 4, pp. 443–464, Aug. 2008.

[12] G. Klein, “The Recognition-Primed Decision (RPD) Model: Looking Back, Looking Forward,” in *Naturalistic Decision Making*, C. E. Zsambok and G. Klein, Eds. Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc, 1997, pp. 285–292.

[13] G. Klein, “Naturalistic Decision Making,” *Hum. Factors*, vol. 50, no. 3, pp. 456–460, Jun. 2008.

[14] D. Kahneman and A. Tversky, “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica.*, vol. 47, no. 2, pp. 263–292, 1979.

[15] H. A. Simon, *Models of Bounded Rationality: Empirically Grounded Economic Reason*. Boston: MIT Press, 1982.

[16] G. Gigerenzer and W. Gaissmaier, “Heuristic Decision Making,” *Annu. Rev. Psychol.*, vol. 62, no. 1, pp. 451–482, Jan. 2011.

[17] R. M. Dawes, D. Faust, and P. E. Meehl, “Clinical versus Actuarial Judgment,” *Science*, vol. 243, no. 4899, pp. 1668–1674, 1989.

[18] P. E. Meehl, “Causes and Effects of My Disturbing Little Book,” *J. Pers. Assess.*, vol. 50, no. 3, pp. 370–375, 1986.

[19] J. F. Nash, *Essays on Game Theory*. Northampton, MA: Edward Elgar Publishing, 1996.

[20] A. Rubinstein, “Comments on the Interpretation of Game Theory,” *Econometrica*, vol. 59, no. 4, pp. 909–924, 1991.

[21] M. J. Osborne and A. Rubinstein, *A Course in Game Theory*. MIT Press, 1994.

[22] W. M. Grove and P. E. Meehl, “Comparative Efficiency of Informal (subjective, impressionistic) and Formal (mechanical, algorithmic) Prediction Procedures: The Clinical–Statistical Controversy.,” *Psychol. Public Policy Law*, vol. 2, no. 2, p. 293, 1996.

[23] F. Faul, E. Erdfelder, A. Buchner, and A.-G. Lang, “Statistical Power Analyses using G*Power 3.1: Tests for Correlation and Regression Analyses,” *Behav. Res. Methods*, vol. 41, no. 4, pp. 1149–1160, Nov. 2009.

[24] de Mel, Geeth, “ISR assets and NIIRS,” personal communication, Jan-2013.

[25] R. H. Hoyle, *Statistical Strategies for Small Sample Research*. Thousand Oaks, CA: SAGE Publications, Inc., 1999.

[26] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. Boca Raton, FL: CRC Press, 1994.

[27] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2013.

[28] Canty, A. and R. Ridpley, *boot: Bootstrap R (S-Plus) Functions*. 2013.

[29] A. Preece, M. Gomez, G. de Mel, W. Vasconcelos, D. Sleeman, S. Colley, G. Pearson, T. Pham, and T. La Porta, “Matching Sensors to Missions Using a Knowledge-Based Approach,” in *Proceedings of SPIE: Defense Transformation and Net-Centric Systems 2008,*, vol. 6981, 2008.

[30] D. Pizzocaro, A. Preece, F. Chen, T. L. Porta, and A. Bar-Noy, “A Distributed Architecture for Heterogeneous Multi Sensor-Task Allocation,” in *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*, 2011, pp. 1–8.

[31] R. Parasuraman and V. Riley, “Humans and Automation: Use, Misuse, Disuse, Abuse,” *Hum. Factors*, vol. 39, no. 2, pp. 230–253, Jun. 1997.

[32] R. Parasuraman and D. H. Manzey, “Complacency and Bias in Human Use of Automation: An Attentional Integration,” *Hum. Factors*, vol. 52, no. 3, pp. 381–410, Oct. 2010.

[33] R. Parasuraman, T. B. Sheridan, and C. D. Wickens, “A Model for Types and Levels of Human Interaction with Automation,” *Syst. Man Cybern. Part A: Syst. Hum., IEEE Trans. on*, vol. 30, no. 3, pp. 286–297, 2000.
[34] N. Silver, *The Signal and the Noise: Why So Many Predictions Fail But Some Don’t*. New York: Penguin, 2012.
[35] “Advanced Chess - Wikipedia, the free encyclopedia.” [Online]. Available: http://en.wikipedia.org/wiki/Advanced_Chess. [Accessed: 30-Jan-2014].
[36] L. G. Terveen, “Overview of Human-Computer Collaboration,” *Knowl.-Based Syst.*, vol. 8, no. 2, pp. 67–81, 1995.
[37] R. J. Crouser and R. Chang, “An Affordance-Based Framework for Human Computation and Human-Computer Collaboration,” *IEEE Trans. Vis. Comput. Graph.*, vol. 18, no. 12, pp. 2859–2868, 2012.
[38] A. B. Wollocko, M. P. Farry, and R. F. Stark, “Supporting Tactical Intelligence Using Collaborative Environments and Social Networking,” in *SPIE Defense, Security, and Sensing*, vol. 8758, 2013
[39] R. Brantingham and A. Hossain, “Crowded: A Crowd-Sourced Perspective of Events as They Happen,” in *SPIE Defense, Security, and Sensing*, vol. 8758, 2013.
[40] A. Preece, C., Gwilliams, C., Parizas, D., Pizzocaro, J.Z., Bakdash, and D., Braines, “Conversational Sensing,” in *Proceedings of SPIE Defense, Security, and Sensing*, 2014.
[41] “Why Her Will Dominate UI Design Even More Than Minority Report | Wired Design | Wired.com.” [Online]. Available: http://www.wired.com/design/2014/01/will-influential-ui-design-minority-report/. [Accessed: 06-Feb-2014].
[42] M. Gomez, A. Preece, M. P. Johnson, G. De Mel, W. Vasconcelos, C. Gibson, A. Bar-Noy, K. Borowiecki, T. La Porta, and D. Pizzocaro, “An Ontology-Centric Approach to Sensor-Mission Assignment,” in *Knowledge Engineering: Practice and Patterns*, Springer, 2008, pp. 347–363.
[43] J. H. Sheehan, P. H. Deitz, B. A. Harris, A. B. Wong, and B. E. Bray, “The Nexus of Military Missions and Means,” DTIC Document, 2004.
[44] D. Pizzocaro, C. Parizas, A. Preece, D. Braines, D. Mott, and J. Z. Z. Bakdash, “CE-SAM: A Conversational Interface for ISR Mission Support,” in *SPIE Defense, Security, and Sensing*,, 2013.
[45] D. Pizzocaro, “Instantaneous Multi-Sensor Task Allocation in Static and Dynamic Environments,” Cardiff University, 2011.
[46] W. R. Shadish, T. D. Cook, and D. T. Campbell, *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin, 2002.
[47] R. E. Nisbett and T. D. Wilson, “Telling More than we can Know: Verbal Reports on Mental Processes.,” *Psychol. Rev.*, vol. 84, no. 3, p. 231, 1977.
[48] F. T. Durso and F. A. Drews, “Health Care, Aviation, and Ecosystems: A Socio-Natural Systems Perspective,” *Curr. Dir. Psychol. Sci.*, vol. 19, no. 2, pp. 71–75, May 2010.
[49] H. S. Kramer and F. A. Drews, “Does the Shoe Fit? Applying Lessons Learned in Aviation to Healthcare,” *Proc. Hum. Factors Ergon. Soc. Annu. Meet.*, vol. 56, no. 1, pp. 936–940, Oct. 2012.

Appendix: Exploratory Results

Additional analyses were performed using descriptive statistics rather than inferential statistics, because there were no a priori hypotheses. Summary statistics with hits, correct rejections, misses, and false alarms for the allocation tasks, collapsed across set, are displayed in Table 5. Misses are omitting the assignment of a sensor capable of performing the task. False alarms are assigning a sensor that is not capable of performing the task.

TABLE 5. DECISION ACCURACY AND ERRORS BY ALLOCATION TASK

	Accuracy		Error	
Allocation Task	**Hits (%)**	**Correct Rejections (%)**	**False Alarms (%)**	**Misses (%)**
1. Moving Car, Jeep, or Humvee	*M* = 79.62 *SD* = 1.89			*M* = 20.38 *SD* = 1.86
2. Moving Military Support Vehicle w/Wheels (such as a Stryker, Transport Truck, or Semi-Truck)	*M* = 79.23 *SD* = 2.03			*M* = 20.77 *SD* = 2.00
3. Known Location, Detect, and Identify a Person with a Hand-Held Missile Launcher	*M* = 27.69 *SD* = 2.03	*M* = 34.23 *SD* = 1.66	*M* = 27.31 *SD* = 1.43	*M* = 10.77 *SD* = 0.01
4. Known Location, Detect and Identify a License Plate on a Vehicle[a]		*M* = 83.85 *SD* = 1.67		*M* = 16.15 *SD* = 1.52
5. Stationary Tank or Other Vehicle w/Tracks	*M* = 83.08 *SD* = 2.04			*M* = 16.92 *SD* = 2.08
6. Deployed Scud Missile Site, Not Covered by Camouflage	*M* = 68.08 *SD* = 3.07			*M* = 31.92 *SD* = 3.03
7. Hole from Digging (1 meter by 1 meter or larger)	*M* = 35.77 *SD* = 1.48	*M* = 26.92 *SD* = 0.01	*M* = 11.54 *SD* = 1.14	*M* = 25.77 *SD* = 1.48
8. Heat from a Running, but Stationary Car	*M* = 33.08 *SD* = 0.01	*M* = 56.54 *SD* = 1.33	*M* = 5.00 *SD* = 2.45	*M* = 5.38 *SD* = 1.13

Empty cells had no responses classified as the respective type of accuracy or error. Standard deviations were calculated by allocation task across participants.

[a.] No ISR platform was capable of reading a license plate. Thus, correct rejection was the only accurate answer.

Mean accuracy (hits and correct rejections) by ISR asset type, collapsed across set and sensor type for brevity, is presented in Table 6.

TABLE 6. DECISION ACCURACY BY ISR ASSET

Asset	Accuracy (%)
Predator A (MQ-1)	*M* = 75.83
Reaper (MQ-9)	*M* = 77.08
Raven	*M* = 76.88
Global Hawk	*M* = 75.21
Shadow 200 (RQ-7)	*M* = 75.00

A summary of free responses to subjective questions is shown in Table 7. Data in Table 7 is presented in a generic aggregate form here and are not shared because some data contains personally identifiable information.

TABLE 7. FREE RESPONSE SUBJECTIVE QUESTIONS

Question[a]	Responses[b]
Set 1: What was your strategy for making these decisions (Examples: *Used your gut or intuition, guessed, etc*)?	*Experience*: 8 out of 10 *No Response*: 1 out of 10 *Some guessing*: 2 out of 10 *Training*: 3 out of 10
Set 2: What was your strategy for making these decisions (Examples: *Looked at NIIRS ratings, guessed, went off of previous decisions, etc*)?	*Experience*: 1 out of 10 *Prior Decisions*: 1 out of 10 *NIIRS ratings*: 8 out of 10 *No Response*: 1 out of 10
General comments about the project?	*Allocation task not representative of the real-world (e.g., operating conditions such as time of day or weather, operator skill, and updates to sensor packages):* 3 out of 10 *No comment:* 7 out of 10 *Video is least operationally valuable type of sensor information*: 1 out of 10

a. Note because some participants had multiple responses the numbers do not sum to 10.

b. Responses categorized into general, paraphrased descriptions.

Tables 8, 9, and 10 summarize responses to the subjective Likert scale questions.

TABLE 8. ISR AND NIIRS

Question[a]	Not at All	A Little Bit	Moderately	Highly	Expert
1. Are you familiar with the capabilities (sensors, speed, etc) of air ISR platforms (such as UAVs)?	0	1	3	3	0
2. Are you familiar with the National Imagery Interpretation Reconnaissance Scale (NIIRS) Ratings for ISR**?**	1	6	2	1	0

TABLE 9. PREVIOUS DECISIONS AND CONFIDENCE RATINGS

Question	No Confidence (Guessing)	Low Confidence	Moderate Confidence	High Confidence	Full Confidence (Certain)
1. Did you use your previous decisions? (referring to using responses in Set 1 for Set 2)	1	2	1	5	1
2. **Set 1:** What is your overall confidence in the sensor assignments to targets?	1	1	5	3	0
3. **Set 2:** What is your overall confidence in the sensor assignments to targets?	0	1	2	3	4

TABLE 10. ISR SYSTEM USEFULLNESS

Question[a]	Never	Once in a While	Sometimes	Often	Always
Would you find a system that recommended or suggested optimal ISR platforms/sensors for target detection tracking helpful?	0	1	2	3	4