THE INTERNATIONAL

C2 JOURNAL

VOLUME 5, NUMBER 1, 2011

SPECIAL ISSUE

Beyond Command and Control: Sense Making under Large World Uncertainty

GUEST EDITOR

Jason K. Levy
Virginia Commonwealth University

The Challenge of iSPIED: intelligence Sense-making to Prognosticate IEDs

Kevin Burns

CCRP
Focus & Convergence for Complex Endeavors

THE INTERNATIONAL C2 JOURNAL

David S. Alberts, Chairman of the Editorial Board, *OASD-NII, CCRP*

The Editorial Board

Éloi Bossé (CAN), *Defence Research and Development Canada*
Berndt Brehmer (SWE), *Swedish National Defence College*
Lorraine Dodd (GBR), *Cranfield University*
Reiner Huber (DEU), *Universitaet der Bundeswehr Muenchen*
William Mitchell (DNK), *Royal Danish Defence College*
Sandeep Mulgund (USA), *The MITRE Corporation*
Mark Nissen (USA), *Naval Postgraduate School*
Mink Spaans (NLD), *TNO Defence, Security and Safety*
Andreas Tolk (USA), *Old Dominion University*

About the Journal

The International C2 Journal was created in 2006 at the urging of an international group of command and control professionals including individuals from academia, industry, government, and the military. The Command and Control Research Program (CCRP, of the U.S. Office of the Assistant Secretary of Defense for Networks and Information Integration, or OASD-NII) responded to this need by bringing together interested professionals to shape the purpose and guide the execution of such a journal. Today, the Journal is overseen by an Editorial Board comprising representatives from many nations.

Opinions, conclusions, and recommendations expressed or implied within are solely those of the authors. They do not necessarily represent the views of the Department of Defense, or any other U.S. Government agency.

Rights and Permissions: All articles published in the International C2 Journal remain the intellectual property of the authors and may not be distributed or sold without the express written consent of the authors.

For more information

Visit us online at: www.dodccrp.org
Contact our staff at: publications@dodccrp.org

CCRP
Focus
& Convergence
for Complex Endeavors

The Challenge of iSPIED: intelligence Sense-making to Prognosticate IEDs

Kevin Burns
(The MITRE Corporation, USA)

Abstract

Sense-making is a cognitive component of command and control, in which analytical inferences are made to support operational decisions. This article examines a prototypical problem in intelligence analysis and addresses the challenge from a psychological perspective. The problem, dubbed iSPIED (intelligence Sense-making to Prognosticate IEDs), reflects a vexing mission of counterinsurgency in Afghanistan, Iraq, and elsewhere in the world. The thesis is that a scientific understanding of sense-making, in countering IEDs or any other threat, requires research grounded in Bayesian theory and formal testing. The result is a design for experimental game tasks, including computational metrics that measure cognitive biases relative to normative standards. The experimental design distinguishes between different dimensions of sense-making, including learning, inference, and choices. The computational metrics are derived from information theory, using entropy to measure bias. Taken together, the design and metrics of iSPIED enable rigorous experiments on cognitive limitations, such as conservatism and confirmation bias, to inform the development of advanced tools and techniques for intelligence analysis.

Introduction

Operational Context

Improvised Explosive Devices (IEDs): *"...continue for the foreseeable future to be the weapon of choice for the world's terrorists, insurgents, militias, guerillas, revolutionaries, and marginal or failed states"* (Zorpette 2008). Billions of dollars have been spent on technological solutions, including better armor and trigger-jamming devices, but with limited success. As a result, the Joint Improvised Explosive Device Defeat Organization (JIEDDO) advocates an approach that includes analytical advancements: *"combining an understanding of the psychology and sociology of terrorist networks with probabilistic modeling, complexity theory, forensic science, pattern recognition, and data mining to predict human behavior..."* (Zorpette 2008).

For example, an approach known as Technosocial Predictive Analytics (AAAI 2009) models the physical and behavioral activities by which IEDs are developed, implanted, and detonated by insurgents. This is done using techniques of quantitative risk assessment (Garrick et al. 2004), to enumerate scenarios and estimate their probabilities and consequences (Apostolakis 1990; Kaplan and Garrick 1980). Often the analyses employ Bayesian Networks (Pearl 2000), which have been demonstrated in principle (Whitney et al. 2009) as being useful for intelligence analysis.

But in practice these predictive methods are limited because the Bayesian Networks require tens or hundreds of probabilities as input, from humans who are biased in various ways (Edwards et al. 1968; Heuer 1999; Kahneman, Slovic, and Tversky 1982; Gilovich, Griffin, and Kahneman 2002). Thus ironically, the Bayesian techniques and tools that have been developed to support human cognition are limited by the same cognitive biases that these methods were developed to overcome in the first place. The biases are still not well understood, although they were recognized half a century ago in the work of Edwards and Phillips (1964) on *"Man as transducer for probabilities in Bayesian command and control systems."* As a result, intelligence analysis continues to be limited by cognitive *sense-making*.

Conceptual Frameworks

Sense-making is central to command and control in countering IEDs (Zorpette 2008), Weapons of Mass Destruction (US Commission on the Prevention of WMD Proliferation and Terrorism 2008), and other threats for which indications and warnings intelligence must anticipate surprise (Grabo 2004). The term "sense-making" itself highlights the psychological and social dimensions of command and control (Leedom 2004), which are distinguished from the informational and physical challenges of planning and execution. According to Alberts and Hayes (2006), *situation awareness* is the primary component of sense-making, and according to Endsley (1988), situation awareness is *"the perception of the elements in the environment within a volume of space and time, the comprehension of their meaning and projection of status in the near future."* Alberts and Hayes (2006) also note that command and control involve sharing of situation awareness, as well as sharing of commanders' intent, in order to achieve desired effects. But before understanding of a situation can be shared between minds, socially, it must first occur within minds, psychologically.

According to Klein et al. (2007), sense-making involves *"reasoning to the best hypothesis"* as follows: *"d is a collection of data. H [hypothesis] explains d. No other hypothesis can explain d as well as H does. Therefore, H is probably true."* But this does not mean that sense-making is only a forensic process, because hypotheses are often concerned with what might occur in the future based on evidence from the past. Indeed the prognostic dimension of sense-making, which corresponds to Endsley's (1988) notion of projection in situation awareness, is especially important for meeting the distributed and dynamic challenges of network centric command and control (Smith 2008;

Alberts, Garstka, and Stein 1999). Also important is the probabilistic dimension of sense-making, i.e., to establish which H *"is probably true."*

Conceptual frameworks like the above, although helpful, are not enough to understand and integrate psychological judgments and computational models in sense-making. A more formal approach would use Bayesian methods to model cognitive sense-making—and thereby gain a deeper understanding of where humans fall short and where humans excel. That approach has been demonstrated for a case study of human error in command and control (Burns 2005b), and a similar approach will be applied here to analytical tasks involved in countering IEDs.

Analytical Gaming

Serious games (Abt 1987) are often played to improve command and control, sometimes with computerized representations of the physical, informational, and social domains. But most of these war games are designed for operational exercises rather than analytical exercises (Ambrose and Ahern 2008; Powers, Stech, and Burns 2010). Also most war games are exercises rather than controlled experiments, and are not intended or instrumented for measuring the cognitive components of command and control.

More rigorous are the lab games used in psychological experiments, such as the Iowa Gambling Task (IGT, see Bechara et al. 1997). IGT is played with four decks of special cards indicating dollar amounts. The decks are face down and the player must make repeated choices, turning the top card of any deck he/she chooses on each trial. A card, when turned, provides a win amount but also incurs a loss amount at probability P. The win/loss amounts and loss probability P are different for each deck, and must be learned by a player as he/she tries to earn the most from repeated choices. By analogy, the player can be likened to a commander who repeatedly sends troops along one of four roads, e.g., to deliver supplies (a win), where each road poses risks of injuries (a loss) from IEDs. Over time, the loss probability P is learned (Burns and Demaree 2009) for each course of action (i.e., choice of an option). Thus there is some similarity, from a cognitive perspective, between playing the IGT and dealing with IEDs.

But what makes these two tasks very different, besides the obvious difference in stakes, is that all knowledge in IGT comes via *episodic learning* from personal experience based on choices—also known as "reinforcement learning" (Doll et al. 2009). In the real-world of command and control, operational knowledge typically comes from intelligence sources rather than personal experience, e.g., from reports of significant activities (SIGACTS) that provide historical data on IED attacks of various types, on various roads, at various times, by various groups, etc. Also, an analyst's challenge is not so much to learn and remember all the instances of *single-source evidence*—but rather to obtain and exploit diverse sources of intelligence, and thereby infer the chances for IED attacks of various types, on various roads, at various times, by various groups, etc. In short, a more relevant game task for intelligence sense-making would pose problems of *probabilistic inference* from *multi-source evidence*.

Mathematical Theory

Inferential Problem

The simplest case of probabilistic inference from multi-source evidence, in H-from-d fashion, would involve only two sources of evidence (d_1 and d_2) and two hypotheses (H_1 and H_2). The prototypical problem of sense-making is to infer the probabilities (P) of hypotheses (H_1 and H_2) in light of evidence (d_1 and d_2), i.e., to determine which H is most probably true. More complex cases would involve more data and hypotheses, but the underlying mathematics are much the same. Here the normative-mathematical approach to solving this problem, known as Bayesian inference (Burns 2005b; 2007), is applied to the cognitive-experimental design of a new game, iSPIED: intelligence Sense-making to Prognosticate IEDs.

In iSPIED, the intelligence analyst's area of interest has North, South, East, and West (N, S, E, and W) roads on which IED attacks can occur, akin to the four decks of cards in IGT. Mathematically, these are four hypotheses (H). The multi-source evidence (d) includes baseline intelligence (e.g., from SIGACTS), called *baseint*, and image intelligence, called *imint*. Importantly, both baseint and imint must be expressed probabilistically for the analyst to infer which H is most probably true. For example, if imint showed that activity was highest on road N, then this might suggest an IED attack is most likely to occur on road N. But without numerical likelihoods provided by imint (and baseint), an analyst would have to make assumptions about likelihoods in order to infer the probabilities of various H given evidence d.

In short, either the data must include likelihoods, or else the analyst must be asked to provide his/her assumed likelihoods, because likelihoods are needed to infer the probabilities of hypotheses from data. Here we focus on the case where players are given baseint and imint in the form of likelihoods. Subsequent experiments might use iSPIED to measure players' assumptions about likelihoods, using data provided in various forms that do not specify likelihoods—and then measure players' multi-source inferences based on their assumptions. But the assumptions are expected to vary widely among people, so a more controlled approach is adopted here to ensure that all participants in experiments have and use the same likelihoods as input to sense-making.

The likelihoods provided to the player are *relative likelihoods* of IED attack on each road, i.e., they are *probabilities* that sum to 100% across the four roads. Each likelihood distribution, across the four roads, is based on a single source of intelligence assuming that it is the only source of intelligence. For example, in iSPIED the baseint (b) likelihoods are based on IED attacks that have been reported in SIGACTS intelligence, whereas the imint (i) likelihoods are based on suspicious actions that have been observed from image intelligence. These two sources of intelligence are independent, and in iSPIED the analyst is also informed that the two sources are equally credible. The analyst's task is to combine the two likelihood distributions in order to infer the relative likelihood of IED attack on each road H (i.e., N, S, E, W) given both sources of data d (i.e., b and i).

The mathematical method for combining likelihoods is given by Bayes Rule, which expresses how a *prior* probability distribution P(H) should be updated using a likelihood distribution p(d|H) to compute a *posterior* probability distribution P(H|d). The equation can be derived from a law of probabilities, written as follows:

$$P(H,d) = P(d) * P(H|d) = P(H) * P(d|H) = P(d,H)$$

where d is data (i.e., from baseint b or imint i) and H is a hypothesis (i.e., N, S, E, or W). Rearranging the middle equality yields Bayes Rule, typically written as follows:

$$P(H|d) = P(H) * P(d|H) / P(d)$$

where P(d) is a normalizing factor given by the sum of terms P(H) * P(d|H) over all H in the frame of discernment. In our case, where the frame of discernment is {N, S, E, W}, we have:

$$P(d) = P(N) * P(d|N) + P(S) * P(d|S) + P(E) * P(d|E) + P(W) * P(d|W)$$

where d denotes either baseint (b) or imint (i).

Thus, in effect, Bayes Rule can be seen as a way to transform probabilities of the form P(d|H) into probabilities of the form P(H|d), using prior knowledge (before getting data d) about P(H). In the special case where one has no prior knowledge, sometimes called an uninformative prior, we have P(N) = P(S) = P(E) = P(W) = 0.25 in our frame of discernment {N, S, E, W}. In this special case, Bayes Rule can be written as follows: $P(H|d) = P(d|H) / P(d|N) + P(d|S) + P(d|E) + P(d|W)$.

As noted earlier, in iSPIED each individual source of intelligence (baseint b or imint i) provides a likelihood distribution P(d|H) across the four hypotheses (N, S, E, W) assuming that d is the only source, which implies a non-informative prior distribution. This allows us to express the likelihoods as *relative* likelihoods for hypotheses (H) given data (d), P(H|d), obtained from P(d|H) by dividing each P(d|H) by the sum $P(d|N) + P(d|S) + P(d|E) + P(d|W)$, as shown above.

These relative likelihoods P(H|d), which are *normalized* to 100%, are provided as the input to sense-making in iSPIED for three reasons. First, normalized probabilities are what iSPIED requires the player to report as output, so it is logical to provide normalized probabilities as input for consistency. Second, a single-source analyst considering multiple hypotheses to explain just one source of evidence would typically report the relative likelihood of each hypothesis given that evidence d, i.e., P(H|d). This in turn would be the input to multi-source inferences, as in iSPIED. Finally, cognitive research (Burns 2006) has found that people naturally think in terms of P(H|d) rather than P(d|H) when faced with the problem of inferring probabilities of hypotheses (H) from data (d). In iSPIED we wish to present likelihoods in a manner that is as logical and typical and natural as possible, to focus on biases in inference rather than biases stemming from the format of input. However one task of iSPIED is designed to measure how cognitive biases may differ when likelihoods are instead presented in the form P(d|H), see Task 5 below.

Now given relative likelihoods of the form P(H|d) as input, the normative process for combining the likelihoods from baseint and imint is again given by Bayes Rule. More specifically, taking the normalized likelihoods $P(H|b)$ for baseint as the prior distribution, the likelihoods $P(i|H)$ for imint can be used to compute the posterior combination of baseint and imint as follows:

$$P(H|b,i) = P(H|b) * P(i|H) \,/\, P(i)$$

where the denominator $P(i)$ is a normalizing factor given by $P(N|b) * P(i|N) + P(S|b) * P(i|S) + P(E|b) * P(i|E) + P(W|b) * P(i|W)$.

Finally, notice that the same result is obtained if likelihoods of the form $P(i|H)$ are replaced by relative (normalized) likelihoods of the form $P(H|i)$, which is the form in which likelihoods are provided as input to iSPIED (as discussed above). Therefore we can write the equation for Bayesian inference from baseint and imint as follows:

$$P(H|b,i) = P(H|b) * P(H|i) \,/\, \Sigma$$

where $H = N, S, E$, or W and $\Sigma = P(N|b) * P(N|i) + P(S|b) * P(S|i) + P(E|b) * P(E|i) + P(W|b) * P(W|i)$.

In words, the Bayesian aggregation of normalized likelihood distributions $P(H|b)$ and $P(H|i)$ is obtained as the normalized product of these two distributions: $P(H|b) * P(H|i) \,/\, \Sigma$. Notice that the product is the same, regardless of whether baseint or imint is treated as the prior.

Psychological Bias

Given the simplicity of the final equation above, one might expect that human beings would be fairly adept at aggregating likelihoods in tasks of Bayesian inference. But in fact research has shown that people exhibit significant biases, even when the task involves only two hypotheses (Burns 2007), let alone for the more complex case of four hypotheses. For example, in various effects known as *Base Rate Neglect* (Tversky and Kahneman 1982), *Representativeness* (Tversky and Kahneman 1974), *Availability* (Tversky and Kahneman 1974), and *Vividness* (Heuer 1999), the first source (baseint) might be discounted or ignored such that the analyst's judgment is biased toward the second source (imint). Conversely, in other effects known as *Anchoring and Adjustment* (Tversky and Kahneman 1974) and *Persistence of Impressions* (Heuer 1999), the second source (imint) might be discounted or ignored such that the analyst's judgment is biased toward the first source (baseint).

Similarly, in *Confirmation Bias* (Lehner et al. 2009; Lehner et al. 2008; Davis 2008; Burns 2005b; Heuer 1999; Klayman and Ha 1987; Fischhoff and Beyth-Marom 1983) an analyst might give more weight to the most likely hypothesis either within the baseint, and/or within the imint, and/or within their product, causing high probabilities to be overestimated and low probabilities to be underestimated. Conversely, in an opposite bias known as *Conservatism*, the analyst's posterior might underestimate high probabilities and overestimate low probabilities, thereby failing to extract all the certainty available in the data (Edwards 1982; Edwards et al. 1968; Phillips and Edwards 1966; Edwards and Phillips 1964).

Although all of these biases have been observed in various experiments, it appears that *Conservatism* is prevalent in tasks of p*robabilistic inference* from *multi-source evidence* (Edwards 1982; Burns 2007). This has practical implications, because often data are expensive or dangerous to obtain, and in those cases it is vital for analysts to extract all the certainty that is warranted. Also of practical concern is that Conservatism is opposite to Confirmation Bias, and Confirmation Bias itself seems to involve at least three component biases that need to be distinguished, as discussed further below.

The biases then get even more complex as more data are aggregated. In iSPIED an additional source of data is signal intelligence, called *sigint*, from electronic surveillance of cell phones and other devices used by enemies to coordinate actions and detonate IEDs. This actually poses two challenges for the player of iSPIED, starting with the choice of a road on which to collect sigint. The choice is a decision-making challenge, for which the normative (optimal) solution can be computed and compared to cognitive responses. Then, after sigint is obtained, the resulting sigint likelihood (s) must be combined with the previous aggregation of baseint and imint (b*i) to get a posterior (b*i*s) that combines all the evidence. But notice that even the earlier problem of decision-making requires Bayesian aggregation of b*i*s for two possible likelihood distributions that may be obtained from sigint, i.e., if signals are found or not found. The reason is that the optimal decision is to collect sigint on whichever road is projected to provide the greatest gain in information, and this depends on the expected change from prior (before sigint) to posterior (after sigint) considering both possibilities (signals found or not found).

To aggregate b*i*s, or any other number of likelihood distributions, Bayes Rule is applied recursively. In the case of three likelihood distributions (b, i, s): First combine baseint and imint (b*i) as discussed above to obtain {P(N|b,i), P(S|b,i), P(E|b,i), P(W|b,i)}. Then, treating this posterior distribution as the new prior distribution, combine it with sigint (s) to compute a new posterior distribution {P(N|b,i,s), P(S|b,i,s), P(E|b,i,s), P(W|b,i,s)}, which is the optimal aggregation of baseint, imint, and sigint.

The various biases noted above may be amplified as the number of data sources is increased to three or more. Biases may also be affected by the order and/or timing and/or format in which data are given, as all of these factors vary in real-world situations. But cognitive biases are not well understood even for the simplest case of combining two sources in a single stage. Also not well understood is how humans acquire individual likelihoods in the first place. Therefore iSPIED presents a series of learning, inference, and choice tasks that become progressively more complex, as shown in Table 1.

Experimental Design

Prototypical Functions

Formal techniques of Cognitive Task Analysis (Bonaceto and Burns 2007; Crandall, Klein, and Hoffman 2006) have been used to perform descriptive studies of sense-making (Pirolli and Card 2005), including *"What makes intelligence analysis difficult?"* (Hutchins, Pirolli, and Card 2007). Similar techniques, including functional decompositions (Elm et al., 2003; Means and Burns 2005), have also been used to design serious games that pose prototypical challenges of command and control (Burns 2005a).

For example, the design of "Pared-down Poker" (Burns 2010) captures a key distinction in command and control (Mandeles, Hone, and Terry 1996) between analytical *inferences* about situations and operational *investments* in courses of action. The former are a challenge of sense-making and the latter are a problem of decision-making. In the case of iSPIED, we are primarily interested in sense-making (see Tasks 1-5 below), but ultimately this sense-making is performed to support decision-making, and often decisions must be made in the process of sense-making. Thus one task of iSPIED (see Task 4 below) requires that the analyst make decisions about where to collect additional intelligence (Pirolli 2007).

Besides the basic difference between sense-making and decision-making, there are functional distinctions to be made between forensic (backward) inference and prognostic (forward) inference. Also important are distinctions discussed earlier between episodic learning from single-source evidence versus probabilistic inference from multi-source evidence, as well as differences between learning of causes and effects versus learning of effects only. These distinctions define different dimensions of sense-making, in five tasks of iSPIED, as outlined in Table 1 and described further below.

Table 1. Cognitive challenges of sense-making, posed by prototypical tasks of iSPIED. Each task is distinguished from the previous task along one dimension (in a column) of interest.

Task 1	Episodic learning of effects	Prognostic	Sense-making
Task 2	Episodic learning of causes and effects	Prognostic	Sense-making
Task 3	Probabilistic inference	Prognostic	Sense-making
Task 4	Probabilistic inference	Prognostic	Sense-making and Decision-making
Task 5	Probabilistic inference	Forensic	Sense-making

Task 1 is a simple task of episodic learning, similar to classic experiments on statistical learning discussed earlier. The stimuli are instances of IED attacks on four roads, with attacks presented one at a time on a graphic display. The player's response, at regular intervals (e.g., once every 20 trials in a task of 100 trials), is a judgment of IED attack probability on each road. This response is measured with a Graphical User Interface (GUI), see Figure 1, which displays probabilistic judgments in a layout similar to the simple map of four roads on which episodic data are presented. For consistency, the same GUI is used to collect probabilistic judgments on all five tasks of iSPIED. Task 1 is useful for measuring biases of memory and recall, such as *Vividness* and *Representativeness*, in episodic learning of likelihoods.

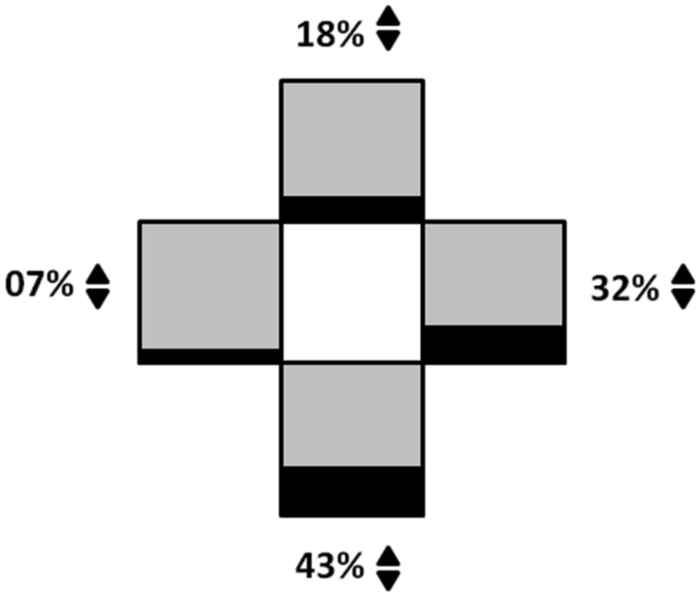

Figure 1. Graphical User Interface (GUI) for collecting human judgments in iSPIED experiments. Each box represents a road (North, South, East, West). Judgments are relative likelihoods of IED attacks, i.e., probabilities that sum to 100%. Judgments are entered numerically, with up and down arrows, and also displayed graphically, as levels in boxes.

Task 2 is a slightly more complex task where the stimuli include both cause and effect. The effects are IED attacks observed on roads, like Task 1, but each report of attack also specifies the group (A, B, C, D) that caused the attack. Like Task 1, the player's response is measured at regular intervals (e.g., once every 100 trials in a task of 300 trials). Unlike Task 1, the player must report not only the probability of IED attack for each road (N, S, E, W) but also the conditional probability of attack by each group (A, B, C, D) on each road. This task is useful for measuring how well players internalize statistical data in a hierarchical structure (Kemp and Tenenbaum 2008; Burns and Demaree 2011) of cause and effect.

Task 3 is a task of probabilistic inference, as opposed to episodic learning in Tasks 1 and 2. This is the central challenge of H-from-d sense-making, which Task 3 poses using four hypotheses and two sources of data. The four hypotheses are roads (N, S, E, W) and the two sources of data are baseint and imint. The stimuli are indications of likelihoods, presented on a graphic display like that used in Tasks 1 and 2. The difference is that here the player is presented with probabilistic evidence (graphically and numerically) rather than episodic instances (graphically and sequentially), and the data are from two independent sources. The task is to combine the likelihoods and report the probability of IED attack on each road. This task will measure biases in Bayesian inference, especially *Conservatism* and *Confirmation Bias*, but also *Base Rate Neglect, Representativeness, Availability, Vividness, Anchoring and Adjustment,* and *Persistence of Impressions*.

Task 4 is similar to Task 3, but provides one more source of data and requires decision-making as well as sense-making. The additional data comes from sigint, and the decision is to choose one road (N, S, E, or W) on which to collect sigint. The player must then use the sigint data to update his/her previous judgment from baseint * imint, and report the resulting probability of IED attack on each road. By presenting three sources of data in a two-stage process, Task 4 will enable further testing of all the same biases noted under Task 3 but extended in a temporal dimension. The decisions in Task 4 will also allow for testing of biases in seeking information. This will help establish if *Confirmation Bias* lies in the *weighing* of likelihoods as they are combined to make inferences, and/or the *seeking* of evidence to support a favored hypothesis. The term "Confirmation Bias" is typically used for both of these component biases. Task 4 includes the seeking dimension as well as the weighing dimension, for comparison to Task 3, which includes only the weighing dimension. Also, even the various components of weighing are not clear, as weighing may arise from episodic *learning* of evidence or probabilistic *inference* from evidence. Tasks 2 and 3 each focus on a different dimension of sense-making, to help establish the relative roles of bias in learning (Task 2) versus bias in inference (Task 3).

Task 5 is a challenge of forensic sense-making about cause and effect. Task 5 presents input likelihoods in the form P(d|H), unlike Tasks 3-4 where likelihoods were presented in the form P(H|d). As discussed above under *Inferential Problem*, both forms are of interest from a psychological perspective although the two forms are equivalent (in the case of iSPIED) from a mathematical perspective. Here the player is presented with three independent likelihood distributions, regarding the size of bomb (small or large), time of day (am or pm), and place on road (left or right) of IED attack. The task is to infer the group that most likely caused the attack, and to report probabilities for each group: P(A), P(B), P(C), and P(D). The inputs are given as likelihoods of effect given cause, e.g., P(size|group), P(time|group), and P(place|group), hence the likelihoods sum to 100% differently than for evidence provided in Tasks 3-4. For example, in Task 5 the likelihoods for P(large|group) might be P(large|A) = 40%, P(large|B) = 20%, P(large|C) = 60%, and P(large|D) = 80%, whereas the likelihoods for P(small|group) might be P(small|A) = 60%, P(small|B) = 80%, P(small|C) = 40%, and P(small|D) = 20%. These likelihoods of the form P(d|H) sum to 100% across each group (H), because all attacks by a given group are either small or large. Thus when told that the IED was small, a player will receive evidence in the form of likelihoods that do not sum to 100%, e.g., for a "small" IED the likelihoods would be {60%, 80%, 40%, 20%} for groups {A, B, C, D}. This is to test if variations in how the problem is formulated and data are presented will affect cognitive biases, compared to Tasks 3 and 4. In another manipulation, some trials of Task 5 present size/time/place likelihoods sequentially in three stages (like the two stages of Task 4), whereas other trials of Task 5 present the identical likelihoods simultaneously. This is to test for biases in attention and limits of working memory, which may also impact sense-making.

Initial Focus

As described above, the five tasks of iSPIED serve to focus initial experiments and analyses on a fundamental understanding of sense-making. But this initial focus is not meant to ignore issues that will clearly require further research on more complex tasks. For example, notice that the *output* of Task 5 is like the *input* to Task 2, i.e., it is a report of the group that caused the IED attack, except that the output of Task 5 is a *probabilistic* report whereas the input to Task 2 is a *deterministic* report. A probabilistic report is more representative of real-world intelligence, where causes are often not known for certain —such that evidence itself is a hypothesis. Future experiments might use iSPIED to study this problem, which reflects the hierarchical nature of real-world intelligence where evidence at one level is a hypothesis at a deeper level (Burns 2005b). The problem is of special concern for team sense-making in collaborative intelligence, where the hypotheses of one analyst in the hierarchy serve as evidence for other analysts up the chain.

The iSPIED tasks are also simplified in another important respect—i.e., the frame of discernment (Burns 2005b) for H-from-d sense-making is *given* to the sense-maker in the form of hypotheses, data, and associated likelihoods. As noted earlier, real-world situations often do not provide likelihoods directly, and in that case any inference will depend on the analyst's *assumptions* about likelihoods for various data in light of various hypotheses. Moreover, and not discussed above, a key task of real-world sense-making is *abduction* (Klein et al. 2007; Thagard 2007; Peirce 1903) to establish the set of hypotheses in the first place—i.e., to create a frame of discernment in which likelihoods can be used to make inferences.

Errors in sense-making may arise from failure to consider a comprehensive set of hypotheses (Heuer 1999), including the hypothesis that sources of data may be deceptive (Stech and Elsässer 2004). Also humans often employ analogies and other abstractions, in the abductions and assumptions that frame their inferences. The present treatment of these matters in iSPIED has been purposely simplified, not to diminish their importance, but rather to enable rigorous scientific research on analytic inferences in initial experiments that control for abductions and assumptions.

Computational Metrics

The five tasks of iSPIED have been programmed in Java software, and human experiments have been performed in a laboratory instrumented for psychological testing. The detailed results of these experiments will be reported elsewhere (see Burns and Demaree 2011). But here it is useful to derive the computational metrics that are needed for the analysis of experimental data and assessment of psychological biases.

The results of experiments and analysis with iSPIED are cognitive-subjective probabilities and normative-objective probabilities, where the latter provide the standard for benchmarking biases in the former. Each point of human data provides a distribution of four probabilities that are directly comparable to Bayesian probabilities. Various comparisons of human-cognitive judgments against Bayesian-normative standards can then be made to answer research questions, discussed above, about how human inferences are biased relative to Bayesian inferences—and how these biases depend on the framing of the problem situation and the nature of the data that are given.

The comparisons require a mathematical measure of difference between cognitive probabilities and normative probabilities. To that end, we denote a set of probabilities as $\{P_f\} = \{P_1, P_2, P_3, P_4\}$ where f is an index that refers to the number of hypotheses (H) in the frame of discernment. Sense-making performance can then be measured by how much certainty the sense-maker has achieved across the set of hypotheses $\{H_f\} = \{H_1, H_2, H_3, H_4\}$. A formal metric is based on the information-theoretic notion of *entropy* (Shannon and Weaver 1949), defined as follows:

$$E = -\Sigma_f \quad P_f * \log_2 P_f.$$

Notice that entropy is highest ($E = 2$) when $\{P_f\} = \{0.25, 0.25, 0.25, 0.25\}$, i.e., when all P_f are equal such that uncertainty across the set of hypotheses is maximal. Conversely, entropy is lowest ($E \approx 0$) when one P_f is ≈ 1.0 and all other P_f in the frame are ≈ 0.0, i.e., when uncertainty across the set of hypotheses is minimal.

Entropy enables us to quantify the normative (optimal) sense-making performance. More specifically we can compute the *negentropy* (N) of $\{P_f\}$ as $N = (E_{max} - E) / E_{max}$, which ranges from 0% to 100% as E ranges from E_{max} to 0. The normative negentropy (denoted N_n) measures how much certainty (fraction of 100%) is achieved in optimal sense-making, so N_n sets the standard for any other sense-maker. By comparison, the cognitive negentropy N_c of a human sense-maker may be more or less than the normative N_n. For example, a human with a *Conservative Bias* will fail to extract all the certainty in the data, so $N_c < N_n$. Conversely, a human with *Confirmation Bias* will overestimate P for the most likely hypothesis and underestimate P for other hypotheses, so $N_c > N_n$. Thus N offers a useful metric for distinguishing between *Conservative Bias* and *Confirmation Bias*, as illustrated in Figure 2.

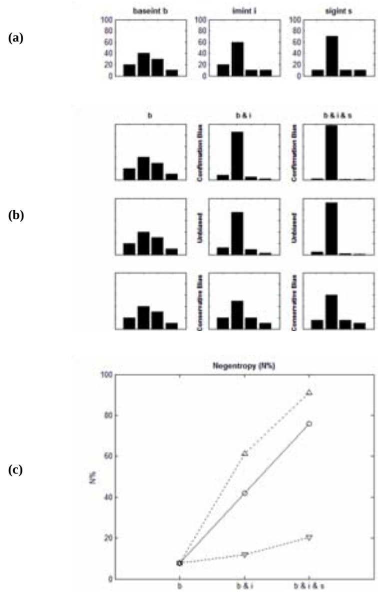

Figure 2. (a) Likelihood distributions across four hypotheses for baseint, imint, and sigint. (b) Posterior probability distributions computed by three models (in three rows) at three stages (in three columns) as more evidence is accumulated. (c) Negentropy of each posterior distribution, plotted at three stages (b; b & i; b & i & s) for three models: up-pointing triangles are Confirmation Bias; circles are unbiased; down-pointing triangles are Conservative Bias.

A related metric goes beyond the *absolute* entropy of a probability distribution $\{P_f\}$ to address the *relative* entropy between two probability distributions $\{P_f\}$ and $\{Q_f\}$. This metric, known as the Kullback-Leibler Divergence (Kullback and Leibler 1951), is defined as follows:

$$K_{pq} = E_{pq} - E_p = -\Sigma_f \; P_f * \log_2 Q_f \; + \; \Sigma_f \; P_f * \log_2 P_f$$

Here K can be seen as the difference between a *cross-entropy* and an entropy, where E_p is the entropy of $\{P_f\}$ and E_{pq} is the cross-entropy of $\{P_f\}$ and $\{Q_f\}$.

Notice that K is zero when $\{P_f\} = \{Q_f\}$, because in that case $E_{pq} = E_p$. Notice also that K increases as the $\{P_f\}$ and $\{Q_f\}$ distributions diverge, regardless of whether E_q is higher or lower than E_p. These features make K a useful measure for the overall *magnitude* of cognitive bias. For example, K_{cn} between the cognitive human and normative standard is a measure of how far the human deviates from the standard. Therefore, K_{cn} (or S_{cn}, discussed below) is a good way to score how well a person has performed in iSPIED. But K_{cn} itself does not capture the *direction* of bias. For example, both a Conservative Bias and Confirmation Bias may lead to the same magnitude of error, hence there is value in also computing negentropy N (discussed above). There is also value in computing other K values, for comparison to K_{cn}, if a descriptive model of bias is available or can be developed.

For example, K_{cd} between the cognitive human and a descriptive model might be compared to two other K values: K_{cn} between cognitive human and normative model; and K_{cu} between cognitive human and uninformative model. Here an uninformative model means a model with maximum entropy: $\{P_f\}_u = \{0.25, 0.25, 0.25, 0.25\}$. The normative model would serve as one null hypothesis (i.e., absence of bias in the probabilities) and the uninformative model would serve as another null hypothesis (i.e., absence of difference in the probabilities) for testing how well the descriptive model captures cognitive biases.

Finally, it is useful to derive a measure of similarity (S) that expresses divergence on a 0-100% scale, like the 0-100% scale of negentropy N. For example, S_{cn} = 100% would be a perfect match between cognitive sense-making and normative sense-making. Because K is a logarithmic measure of information, an inverse log gives us the following expression for S:

$$S = 100\% * 2^{-K}.$$

Figures 3 and 4 illustrate two examples, comparing S_{cn}, S_{cd}, and S_{cu}. When $S_{cn} < S_{cd} > S_{cu}$, as in Figure 3, then the descriptive model is a good model because it captures cognitive biases better than a normative model and better than an uninformative model. When $S_{cn} > S_{cd} < S_{cu}$, as in Figure 4, then the descriptive model is a poor model of cognitive sense-making.

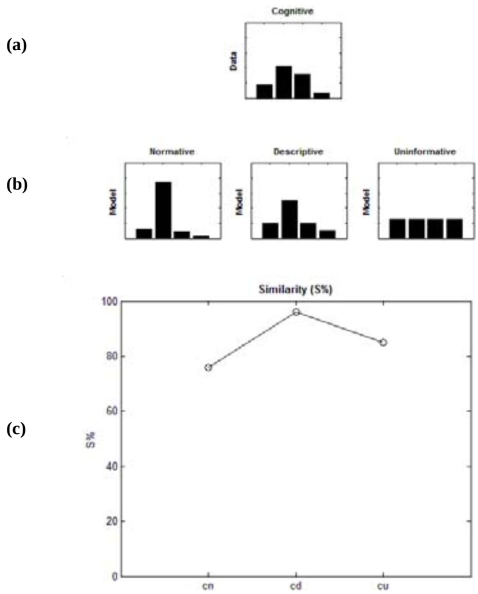

Figure 3. (a) Probability Distribution for Cognitive data. (b) Probability Distributions for three models: (n) Normative, (d) Descriptive, (u) Uninformative. (c) Similarity (S%) of Cognitive data to three different models: S_{cn}, S_{cd}, S_{cu}. Here the descriptive model is a good model because $S_{cn} < S_{cd} > S_{cu}$.

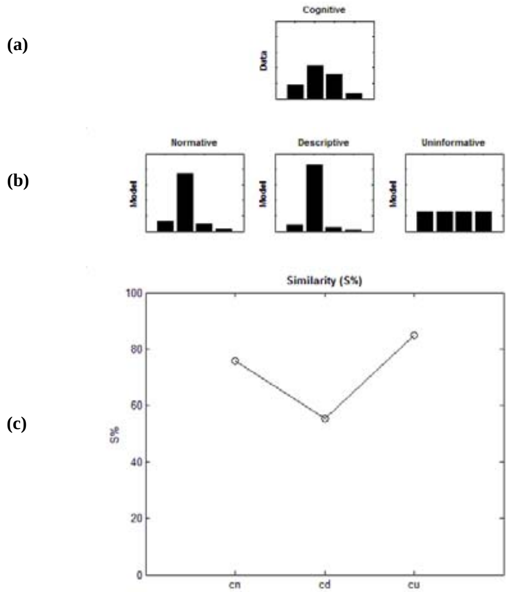

Figure 4. *(Refer to Figure 3)* **Here the descriptive model is a poor model because $S_{cn} > S_{cd} < S_{cu}$.**

As highlighted by S_{cn}, cognitive data from experiments must be measured with respect to some normative standard in order to establish sense-making performance. This is not the same as comparing to ground truth, because optimal sense-making can only be as effective as the Bayesian inference from available evidence. Similarly, sense-making cannot be measured simply by comparing a sense-maker's decision to the normative decision in a forced choice task. For example, when forced to choose the best road to take (e.g., to avoid IED attacks), the resulting choice will tell us very little about the underlying probabilities that are the product of sense-making and inputs to decision-making (Burns and Demaree 2009). Therefore, compared to other metrics like ground truth or other methods like forced choice, the information-theoretic approach given by N, K, and S above offers advantages for measuring and modeling sense-making.

Practical Insights

The main contribution of this article is an *Experimental Design*, along with underlying *Mathematical Theory* and associated *Computational Metrics*. The design, theory, and metrics offer several practical insights, especially regarding the importance of *cognitive task distinctions, conditional likelihoods,* and *measures of biases.*

With respect to cognitive task distinctions, the challenges of sense-making have previously been described by others informally in *Conceptual Frameworks*. iSPIED goes further to formally distinguish between different functions of cognition, especially: sense-making versus decision-making; prognostic inference versus forensic inference; and probabilistic inference versus episodic learning. These distinctions are crucial for rigorous understanding of cognitive sense-making and related processes, so that the relevant processes can be measured and modeled via psychological experiments and computational analyses.

With respect to conditional likelihoods, an important insight from iSPIED is that any inference (and especially the optimal Bayesian aggregation of likelihoods) requires likelihoods as input in the first place. Although this may seem somewhat obvious, the central role of likelihoods has not been recognized in most *Conceptual Frameworks* proposed by others. For example, authors of the data-frame theory acknowledge that the purpose of sense-making is to infer which *"H [hypothesis] is probably true"* (Klein et al. 2007, 125). But these authors never mention likelihoods, and do not address how likelihoods would be learned and used to establish which H is *probably* true, e.g., via probabilities in Bayesian inference. Instead they describe a frame more vaguely as *"an explanatory structure that defines entities by describing their relationship to other entities... [which] can take the form of a story... a map... a script... or a plan"* (118). The same authors also write that *"a frame is a structure for accounting for the data and guiding the search for more data"* (125).

The design of iSPIED serves to formalize the various entities distinguished as hypotheses (H) and evidence (d), and to specify the knowledge structures needed for sense-making—which are *likelihoods* of data (d) given hypotheses (H), i.e., $P(d|H)$. A Bayesian framework specifies how a posterior probability of the form $P(H|d)$ can be obtained by combining a likelihood $P(d|H)$ with a prior $P(H)$, and thereby infer which H is most probably true given d. The same Bayesian framework also specifies how additional data can be applied to update the probabilities of hypotheses, in multi-source inferences, i.e., by recursive application of Bayes Rule.

In short, data are useful for sense-making only when accompanied by the necessary knowledge structures—and the necessary knowledge structures are not stories or maps or scripts or plans but rather *likelihoods* of the form $P(d|H)$ and $P(H|d)$. Although these likelihoods may be implicit in stories or other notions of frames, the likelihoods must be made explicit in any formal account of sense-making—as noted in early research on *"Man as Transducer for Probability in Bayesian Command and Control Systems"* (Edwards and Phillips 1964). In that work it was proposed humans could be used to estimate likelihoods of the form $P(d|H)$, but machines would be needed to aggregate the likelihoods across multiple data sources to compute $P(H|d_1,d_2,d_3,...)$— i.e., because humans were found to be biased in the aggregation task.

A key insight from iSPIED is that estimation of probabilities and aggregation of those probabilities may not be so easily separated, even in a simple game let alone the real-world. The reason is that both tasks (estimation and aggregation) are governed by cognitive heuristics and biases in probabilistic reasoning. Even initial estimates of probabilities (later to be aggregated) often require the combination of data presented episodically and/or probabilistically, so an integrated approach will be needed to understand the cognitive processes and uncover the relevant biases. iSPIED takes a step in this direction with a suite of tasks, including Tasks 1-2 that focus on learning and estimation of single-source probabilities, as well as Tasks 3-5 that focus on inference in aggregation of multi-source probabilities.

The formal framework of iSPIED can also shed light on the informal notion often expressed as *"connecting the dots."* This phrase is superficial (see Lowenthal 2008) and perhaps even detrimental because it obscures what dots represent and how dots can and should be connected. Typically the term "dots" implies only data (d), but in iSPIED we see that the dots must include hypotheses H as well as data d in order to accomplish H-from-d sense-making. Moreover, iSPIED shows that sense-making requires not only d and H but also the conditional *likelihoods* that relate data to hypotheses.

Recent critiques of intelligence practices have focused on sharing data between (and within) agencies, in the aftermath of failures like those of 9/11, Iraq's WMDs, etc. Yet the key to effective intelligence lies not in sharing data d but rather in knowledge of likelihoods that are needed for H-from-d sense-making. The relative likelihoods of various hypotheses, given a collection of evidence, can be computed only if likelihoods for each individual datum d and hypothesis H are known (and then aggregated). Sharing data is useful only if the associated knowledge of likelihoods is also shared, or if receivers of the data are capable of accurately estimating the likelihoods themselves. Bias and errors can arise from sharing data in group sense-making, if the receivers are not able to accurately estimate likelihoods for all competing hypotheses that they should consider in their frame of discernment (Heuer 1999).

Moving beyond the notion of dots, a *"signal-noise"* analogy from Signal Detection Theory (SDT) has also been proposed for analyzing intelligence failures, e.g., in Wohlstetter's (1962) well-known study of Pearl Harbor. Her claim is that the US was surprised by the attack because we failed to identify signals of impending attack in the context of non-signals (noise). A problem with SDT, however, is that it only makes binary distinctions between a signal (H) and non-signals (~H). Even in simple cases and especially in complex cases like Pearl Harbor, sense-makers must almost always consider more than two hypotheses in their frame of discernment. For example, in the context of an insurgency in which IEDs are a concern, a binary distinction between friendly (H) and not-friendly (~H) will fail to distinguish between important classes of not-friendly populations, such as neutral versus hostile.

Finally, with respect to measures of biases, it is important to acknowledge that retrospective analysis of any case study will be limited by a lack of details about the actual likelihoods and associated data that were available to sense-makers at the time they needed to make sense. As such, analysis by anecdote (Medina 2008) using case studies (Davies 2008) cannot establish the existence of cognitive biases, scientifically, let alone their magnitudes and directions, systematically. Instead the biases must be measured empirically and modeled computationally, using test beds like iSPIED, in order to gain a more rigorous understanding of sense-making.

This understanding, in turn, is needed to design advanced tools and techniques for improving intelligence analysis. Today's tools for intelligence analysis serve many useful functions, such as searching large databases and generating visualizations. But these tools do little to address the cognitive challenges of sense-making, which involve estimating and aggregating likelihoods of various data d and hypotheses H in frames of discernment. One exception comes from the technique known as Analysis of Competing Hypotheses (Heuer 1999), and associated tools (Heuer 2008), which are designed to help mitigate Confirmation Bias by assessing evidence across a comprehensive frame of discernment. Another example is the graphic system known as Bayesian Boxes, which was developed to improve inferences via "structure mapping" visualizations of likelihoods, priors, and posteriors (Burns 2007; 2006). Additional tools and techniques might also be invented and improved, with proper grounding in computational theory, psychological models, and empirical testing.

Conclusion

This article began by introducing analytical challenges involved in countering IEDs, and identified shortcomings in the current state of *Conceptual Frameworks* and *Analytical Gaming* used to study sense-making. The main contributions were then presented in a *Mathematical Theory* addressing *Psychological Bias*, along with an *Experimental Design* including *Computational Metrics* for formally assessing the magnitudes and directions of cognitive biases.

The value of iSPIED lies in its design, theory, and metrics, which together are needed to advance a scientific understanding of sense-making. This science, in turn, is needed to develop practical applications in the form of training and tools that can improve sense-making. The problems of bias in tasks of Bayesian inference have been known for decades, based on basic research (Edwards and Phillips 1964; Edwards et al. 1968) as well as applied studies (Zlotnick 1970; Fisk 1972; Schweitzer 1976). However, recent research on sense-making has typically been more conceptual (Klein et al. 2007; Alberts and Hayes 2006; Leedom 2004; Fishbein and Treverton 2004; Weick and Sutcliffe 2001) and anecdotal (Davis 2008; Medina 2008). A computational approach supported by empirical testing, as in iSPIED, is needed to deepen our understanding of sense-making.

A computational approach is especially valuable for bridging the gap between machine systems and human users in human-system integration. Systems such as Bayesian Networks, discussed in the introduction to this article, will be of limited use unless they are designed to work in concert with the humans that might employ them. Moreover, cognitive heuristics are often efficient (Gigerenzer and Todd 1999), and in that case worthy of being implemented in computer systems that could realize the same efficiencies. But to do so, the cognitive heuristics must first be modeled computationally. iSPIED provides a test bed for measuring and modeling these heuristics and biases, in sense-making tasks that present cognitive challenges prototypical of multi-source intelligence analysis.

References

Association for the Advancement of Artificial Intelligence (AAAI). 2009. Spring Symposium on *Technosocial Predictive Analytics*, March 23-25, Stanford University, Stanford CA.

Abt, C. 1987. *Serious Games*. Lanham, MD: University Press of America.

Alberts, D., and R. Hayes. 2006. *Understanding Command and Control.* Washington, DC: Command and Control Research Program (CCRP) Publications.

Alberts, D., J. Garstka, and F. Stein. 1999. *Network Centric Warfare: Developing and Leveraging Information Superiority*, 2nd ed. Washington, DC: CCRP Publications.

Ambrose, F., and B. Ahern. 2008. Unconventional Red Teaming. In *Anticipating Rare Events: Can Acts of Terror, Use of Weapons of Mass Destruction, or Other High Profile Acts Be Anticipated?*, ed. N. Chesser, 136-139. <http://redteamjournal.com/papers/U_White_Paper-Anticipating_Rare_Events_Nov2008rev.pdf>

Apostolakis, G. 1990. The Concept of Probability in Safety Assessments of Technological Systems. *Science* 250: 1359-1364.

Bechara, A., H. Damasio, D. Tranel, and A. Damasio. 1997. Deciding Advantageously Before Knowing the Advantageous Strategy. *Science* 275: 1293-1295.

Bonaceto, C., and K. Burns. 2007. A Survey of the Methods and Uses of Cognitive Engineering. In *Expertise Out of Context,* ed. R. Hoffman, 29-75. New York, NY: Lawrence Erlbaum.

Burns, K. 2010. Strategic Style in Pared-down Poker. In *The Structure of Style: Algorithmic Approaches to Manner and Meaning,* ed. S. Argamon, K. Burns, and S. Dubnov, 247-289. Springer-Verlag.

Burns, K. 2007. Dealing with Probabilities: On Improving Inferences with Bayesian Boxes. In *Expertise Out of Context: Proceedings of the Sixth International Conference on Naturalistic Decision Making,* ed. R. Hoffman, 263-280. New York, NY: Lawrence Erlbaum.

Burns, K. 2006. Bayesian Inference in Disputed Authorship: A Case Study of Cognitive Errors and a New System for Decision Support. *Information Sciences* 176: 1570-1589.

Burns, K. 2005a. Pared-down Poker: Cutting to the Core of Command and Control. Proceedings of the *IEEE Conference on Computational Intelligence in Games*: 234-241.

Burns, K. 2005b. Mental Models and Normal Errors. In *How Professionals Make Decisions,* ed. H. Montgomery, R. Lipshitz, and B. Brehmer, 15-28. Mahwah, NJ: Lawrence Erlbaum.

Burns, K., and H. Demaree. 2011. How Humans Forget Frequencies: Psychological Limits in Statistical Learning of Hierarchical Likelihoods. *Unpublished manuscript.*

Burns, K., and H. Demaree. 2009. A Chance to Learn: On Matching Probabilities to Optimize Utilities. *Information Sciences* 179(11): 1599-1607.

Crandall, B., G. Klein, and R. Hoffman. 2006. *Working Minds: A Practitioner's Guide to Cognitive Task Analysis.* Cambridge, MA: MIT Press.

Davis, J. 2008. Why Bad Things Happen to Good Analysts. In *Analyzing Intelligence: Origins, Obstacles, and Innovations,* ed. R. George and J. Bruce, 157-170. Washington, DC: Georgetown University Press.

Doll, B., W. Jacobs, A. Sanfey, and M. Frank. 2009. Instructional Control of Reinforcement Learning: A Behavioral and Neurocomputational Investigation. *Brain Research* 1299: 74-94.

Edwards, W. 1982. Conservatism in Human Information Processing. In *Judgment Under Uncertainty: Heuristics and Biases,* ed. D. Kahneman, P. Slovic, and A. Tversky, 359-369. Cambridge, UK: Cambridge University Press.

Edwards, W., L. Phillips, W. Hayes, and B. Goodman. 1968. Probabilistic Information Processing Systems: Design and Evaluation. *IEEE Transactions on Systems, Man, and Cybernetics* 4(3): 248-265.

Edwards, W., and L. Phillips. 1964. Man as Transducer for Probabilities in Bayesian Command and Control Systems. In *Human Judgments and Optimality,* ed. M. Shelly and G. Bryan, 360-401. New York, NY: Wiley & Sons.

Elm, W., S. Potter, J. Gualtieri, J. Easter, and E. Roth. 2003. Applied Cognitive Work Analysis: A Pragmatic Methodology for Designing Revolutionary Cognitive Affordances. In *Handbook of Cognitive Task Design,* ed. E. Hollnagel, 357-382. Mahwah, NJ: Lawrence Erlbaum.

Endsley, M. 1988. Design and Evaluation for Situation Awareness Enhancements. Proceedings of the *32nd Annual Meeting of the Human Factors Society*: 97-101.

Fischhoff, B., and R. Beyth-Marom. 1983. Hypothesis Evaluation from a Bayesian Perspective. *Psychological Review* 90(3): 239-260.

Fishbein, W., and G. Treverton. 2004. Rethinking "Alternative Analysis" to Address Transnational Threats. *Occasional Papers* 3(2), Sherman Kent Center for Intelligence Analysis.

Fisk, C. 1972. The Sino-Soviet Border Dispute: A Comparison of the Conventional and Bayesian Methods for Intelligence Warning. *Studies in Intelligence* 16(2): 53-62.

Garrick, B., J. Hall, M. Kilger, J. McDonald, T. O'Toole, P. Probst, E. Parker, R. Rosenthal, A. Trivelpiece, L. Van Arsdale, and E. Zebroski. 2004. Confronting the Risks of Terrorism: Making the Right Decisions. *Reliability Engineering and System Safety* 86(2): 129-176.

Gigerenzer, G., and P. Todd. 1999. *Simple Heuristics That Make Us Smart.* Oxford, UK: Oxford University Press.

Gilovich, T., D. Griffin, and D. Kahneman. 2002. *Heuristics and Biases: The Psychology of Intuitive Judgment.* Cambridge, UK: Cambridge University Press.

Grabo, C. 2004. *Anticipating Surprise: Analysis for Strategic Warning.* Lanham, MD: University Press of America.

Heuer, R. 2008. Computer-aided Analysis of Competing Hypotheses. In *Analyzing Intelligence: Origins, Obstacles, and Innovation*, ed. R. George and J. Bruce, 251-265. Washington, DC: Georgetown University Press.

Heuer, R. 1999. *Psychology of Intelligence Analysis*. Washington, DC: Center for the Study of Intelligence.

Hutchins, S., P. Pirolli, and S. Card. 2007. What Makes Intelligence Analysis Difficult? A Cognitive Task Analysis. In *Expertise Out of Context,* ed. R. Hoffman, 281-316. New York, NY: Lawrence Erlbaum.

Kahneman, D., P. Slovic, and A. Tversky. 1982. *Judgment Under Uncertainty: Heuristics and Biases*. Cambridge, UK: Cambridge University Press.

Kaplan, S., and B. Garrick. 1980. On the Quantitative Definition of Risk. *Risk Analysis* 1(1): 11-27.

Kemp, C., and J. Tenenbaum. 2008. The Discovery of Structural Form. Proceedings of the *National Academy of Sciences* 105(31): 10687-10692.

Klayman, J., and Y. Ha. 1987. Confirmation, Disconfirmation, and Information in Hypothesis Testing. *Psychological Review* 94(2): 211-228.

Klein, G., J. Philips, E. Rall, and D. Peluso. 2007. A Data-frame Theory of Sensemaking. In *Expertise Out of Context*, ed. R. Hoffman, 113-155. New York, NY: Lawrence Erlbaum.

Kullback, S., and R. Leibler. 1951. On Information and Sufficiency. *Annals of Mathematical Statistics* 22: 79-86.

Leedom, D. 2004. The Analytic Representation of Sensemaking and Knowledge Management within a Military C2 Organization. *United States Air Force Research Laboratory Technical Report. AFRL-HE-WP-TR-2004-0083*. Wright-Patterson Air Force Base, OH: <http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA428100&Location=U2&doc=GetTRDoc.pdf>

Lehner, P., L. Adelman, R. DiStacio, M. Erie, J. Mittel, and S. Olson. 2009. Confirmation Bias in the Analysis of Remote Sensing Data. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* 39(1): 218-226.

Lehner, P., L. Adelman, B. Cheikes, and M. Brown. 2008. Confirmation Bias in Complex Analyses. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans* 38(3): 584-592.

Lowenthal, M. 2008. Intelligence in Transition: Analysis after September 11 and Iraq. In *Analyzing Intelligence: Origins, Obstacles, and Innovations*, ed. R. George and J. Bruce, 226-237. Washington, DC: Georgetown University Press.

Mandeles, M., T. Hone, and S. Terry. 1996. *Managing Command and Control in the Persian Gulf War.* Westport, CT: Praeger.

Means, D., and K. Burns. 2005. Analyzing Decisions and Characterizing Information in C2 Systems. Paper presented at the *10th International Command and Control Research and Technology Symposium*, June 13-16, McLean, VA.

Medina, C. 2008. The New Analysis. In *Analyzing Intelligence: Origins, Obstacles, and Innovations*, ed. R. George and J. Bruce, 238-249. Washington, DC: Georgetown University Press.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

Peirce, C. 1903. C. S. Peirce's Lowell Lectures of 1903. Eighth Lecture: *Abduction*. MS 475: 2-92.

Phillips, L,. and W. Edwards. 1966. Conservatism in a Simple Probability Inference Task. *Journal of Experimental Psychology* 72: 346-357.

Pirolli, P. 2007. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford, UK: Oxford University Press.

Pirolli, P., and S. Card. 2005. The Sensemaking Process and Leverage Points for Analyst Technology as Identified Through Cognitive Task Analysis. Proceedings of the *International Conference on Intelligence Analysis, IA '05*, May 2-4, McLean, VA.

Powers, E., F. Stech, and K. Burns. 2010. A Behavioral Model of Team Sensemaking. *International Command and Control Journal* 4(1): 1-10.

Schweitzer, N. 1976. Bayesian Analysis for Intelligence: Some Focus on the Middle East. *Studies in Intelligence* 20(2): 31-44.

Shannon, C., and W. Weaver. 1949. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Smith, T. 2008. Predictive Warning: Teams, Networks, and Scientific Method. In *Analyzing Intelligence: Origins, Obstacles, and Innovations*, ed. R. George and J. Bruce, 266-280. Washington, DC: Georgetown University Press.

Stech, F., and C. Elsässser. 2004. Midway Revisited: Detecting Deception by Analysis of Competing Hypotheses. Paper presented at *72nd MORS Symposium,* June 22-24, Naval Postgraduate School, Monterey, CA.

Thagard, P. 2007. Abductive Inference: From Philosophical Analysis to Neural Mechanisms. In *Inductive Reasoning: Experimental, Developmental, and Computational Approaches,* ed. A. Feeney and E. Heit, 226-247. Cambridge, UK: Cambridge University Press.

Tversky, A., and D. Kahneman. 1982. Evidential Impact of Base Rates. In *Judgment Under Uncertainty: Heuristics and Biases,* ed. D. Kahneman, P. Slovic, and A. Tversky, 153-160. Cambridge, UK: Cambridge University Press.

Tversky, A., and D. Kahneman. 1974. Judgment Under Uncertainty: Heuristics and Biases. *Science* 185: 1124-1131.

United States Commission on the Prevention of Weapons of Mass Destruction Proliferation and Terrorism: B. Graham, J. Talent, G. Allison, R. Cleveland, S. Rademaker, T. Roemer, W. Sherman, H. Sokolski, and R. Verma. 2008. *World at Risk: The Report of the Commission on the Prevention of WMD Proliferation and Terrorism*. New York, NY: Vintage Books.

Weick, K., and K. Sutcliffe. 2001. *Managing the Unexpected*. San Francisco, CA: Jossey-Bass.

Whitney, P., A. Brothers, G. Coles, Y. Young, K. Wolf, S. Thompson, D. Niesen, J. Madsen, and C. Henderson. 2009. Technosocial Modeling of IED Threat Scenario and Attack. Proceedings of *AAAI Symposium on Technosocial Predictive Analysis*: 142-147.

Wohlstetter, R. 1962. *Pearl Harbor: Warning and Decision*. Stanford, CA: Stanford University Press.

Zlotnick, J. 1970. Bayes' Theorem for Intelligence Analysis. Paper presented at the *Conference on the Diagnostic Process*, June 18, Ann Arbor, MI.

Zorpette, G. 2008. Countering IEDs. *IEEE Spectrum* September: 26-35.