Understanding Sensemaking Using Functional Architectures

Robert Thomson, Christian Lebiere, Matthew Rutledge-Taylor, James Staszewski, and John R. Anderson
Carnegie Mellon University
5000 Forbes Avenue
Pittsburgh, PA 15218
412-268-6284
thomsonr@andrew.cmu.edu, cl@andrew.cmu.edu, mattrt@andrew.cmu.edu, jjs@cmu.edu, ja+@cmu.edu

Keywords:
Sensemaking, Biases, Functional Modeling, Cognitive Architectures

ABSTRACT: *This paper outlines similarities between sensemaking theory and the ACT-R cognitive architecture. We analyze a functional model that interprets geospatial imagery data implemented in the ACT-R cognitive architecture. We also discuss how the various cognitive mechanisms of the functional model fit within sensemaking theory, and finally how an analysis of these mechanisms may give rise to cognitive biases.*

1. Introduction

When people make decisions, they must gather, elaborate, distill, and process (potentially) incomplete, incorrect, or contradictory information from the environment into actionable decisions. Sensemaking is a qualitative description of how information is gathered, structured, and used to generate and revise hypotheses (see Figure 1.1). Information flows through two interconnected processing loops: the foraging loop and the sensemaking loop (Pirolli & Card, 2005).

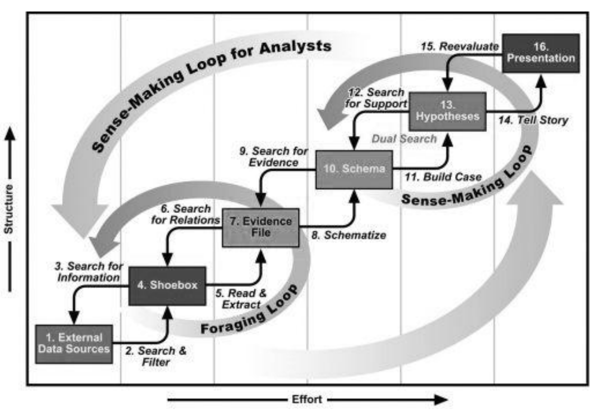

Figure 1.1. An overview of sensemaking. Reproduced from Interactive Automation. Retrieved January 31, 2012, from http://dydan.rutgers.edu/PDDALab/dev/images/flow.png.

Foraging describes how data is gathered, filtered, and aggregated into structured evidence. It is a form of (generally) bottom-up data collection. The stages in the foraging loop can be abstracted to a data gathering and implicit learning process that filters and assimilates information (Kahneman & Treisman, 1984).

The sensemaking loop describes how hypotheses are generated, evaluated, and either revised or discarded. Hypotheses drive top-down processes such as guiding attention to relevant information through the application and interpretation of schemas. Schemas are knowledge structures that both organize data and shape how this data is interpreted. For instance, the schema for a house fire is different whether you are the homeowner, the firefighter, or the arson investigator (Klein, Moon, & Hoffman, 2006). The concept of a frame, used by Klein, Phillips, Rall & Peluso (2006), is essentially equivalent to the sensemaking concept of a schema. In accordance with the fuller body of sensemaking literature, we prefer to use the term frame.

In the sensemaking loop, hypotheses are either revised or discarded when conflicting data enters the system (e.g., when evidence contrary to a hypothesis is encountered). This generally results from a misclassification of data within a frame. The hypothesis (and frame) needs to be revised to fit the new data, or a new hypothesis has to be adopted. Cognitive processes such as insight learning and analogical reasoning are generally incorporated into explanations of the sensemaking loop.

From a cognitive perspective, sensemaking can be broken into six main processes: learning a frame, recalling a frame, assessing the current frame, generating hypotheses, acquiring additional data, and reframing based on this evidence. We will focus on these six component sensemaking processes and how they can be mapped to mechanisms within the ACT-R cognitive architecture.

ACT-R 6 is a computational implementation of a unified theory of cognition. It accounts for information processing in the mind via task-invariant mechanisms constrained by the biological limitations of the brain. While sensemaking theory abstracts away from brain processes, it makes commitments to the control and flow of information that are commensurable with ACT-R's functional perspective. For example, the processing loops in sensemaking can be instantiated in the production rules controlling the follow of control and information in ACT-R. Furthermore, ACT-R is committed to localization of neural architecture, allowing for functional models to guide the development of neurally-inspired models.

To describe the mapping of sensemaking onto ACT-R, we will describe two models of very different tasks from the IARPA-funded ICArUS-MINDS project. Its goal is to create a neurally plausible model of sensemaking that accounts for cognitive biases in the context of intelligence analysis.

1.1 The ACT-R 6 Architecture

ACT-R is a functional cognitive architecture used to model diverse cognitive phenomena. The ACT-R architecture includes long-term declarative memory, procedural memory, and perceptual-motor modules connected through limited-capacity buffers. When a retrieval request is made to declarative memory (DM), the most active matching chunk is returned, where activation is computed as the sum of base-level activation, spreading activation, mismatch penalty and stochastic noise.

Spreading activation is a mechanism that propagates activation from the contents of buffers to declarative memory proportionally to their strength of association. Partial matching is a mechanism that allows for chunks in memory that do not perfectly match a retrieval request to be recalled if their activation overcomes a similarity-based mismatch penalty. Blending is a mechanism similar to partial matching that allows for a memory retrieval that results in a new chunk being created that reflects the consensus of all chunks in memory proportional to their activation instead of the retrieval of an existing chunk.

The flow of information is controlled in ACT-R by a production system, which operates on the contents of the buffers. Each production consists of if-then condition-action pairs. Conditions are typically criteria for buffer matches, while the actions are typically changes to the contents of buffers that might trigger operations in the associated modules. The production with the highest utility is selected to fire from among the eligible productions. Please see Anderson and Lebiere (1998) and Anderson et al. (2004) for a more complete account of the mechanisms implemented in the ACT-R architecture.

2. Sensemaking and ACT-R

ACT-R has previously been used to model several basic components of sensemaking such as evidence marshaling (Pirolli, Fu, Reeder & Card, 2002), and categorization (Anderson & Betz, 2001). The perceptual-motor modules and imaginal buffer in ACT-R are architectural analogues of the foraging processes in sensemaking theory. The visual buffer is used to store a low-level representation of visual stimuli in the environment. The imaginal buffer can be used for the creation of new chunks which are then stored into declarative memory.

In sensemaking, the foraging loop is a process of perceiving information from an external data source, placing it into a catch-all 'shoebox', and then organizing this 'raw' information into a series of structured evidence files. ACT-R's visual module and imaginal buffer have functionality that can be used to mirror these information-gathering (i.e., foraging) steps in sensemaking theory. Specifically, the visual buffer 'perceives' spatial and object information (from the visicon; ACT-R's representation of an external data source). This information is then harvested in a 'raw' form into declarative memory, which is analogous to the 'shoebox'. Productions, however, can also aggregate this raw data and place it in the imaginal buffer, which then creates a new organized chunk in declarative memory for later retrieval (i.e., an evidence file). Based on the representational complexity of the task it may be necessary to aggregate perceptual information using the imaginal buffer. For instance, if the raw perceptual representation contains more information than is required for the task, a representation that focuses on the relevant features can be created. Doing so would enhance architectural mechanisms such as spreading activation and improve the efficiency of the learning process.

As previously discussed, the sensemaking loop contains six main cognitive processes: learning, recalling, and assessing a frame, generating hypotheses, acquiring additional data, and reframing. While these processes are represented as separate steps in the sensemaking loop, the cognitive processes subsuming their function are not necessarily distinct. In ACT-R, both frames and hypotheses can be represented as chunks. In general, the difference between frames and hypotheses is the kind of information stored in the chunks, which buffer holds the chunk (e.g., the goal buffer holding hypotheses, and the retrieval or imaginal buffers holding frames), and how productions manipulate the chunk structure.

Framing (learning, recalling and assessing a frame) involves recalling information that was encoded in the foraging loop and then applying an organizational perspective to it. In ACT-R, a frame can be represented by a chunk or multiple related chunks holding rule-like information for the organizing and interpreting of data-chunks (i.e., evidence files) into testable hypotheses. Based on task complexity, it may not be necessary to represent a frame as a series of related (generally hierarchically-organized) chunks if the expected output can be captured in a simple rule-like structure.

In many sensemaking tasks, hypotheses take the form of either an estimate of a forced-choice response or the generation of likelihoods of the presence or absence of a given state of the environment. In ACT-R, a hypothesis can be represented as a chunk that contains the representation of a potential response. An initial hypothesis allows for the model to test against either an actual or theoretical outcome and guides the gathering of additional evidence that leads to reframing.

Gathering more data and reframing occur through feedback on the accuracy of hypotheses and by intuiting regularities in new data gathered by the foraging loop. Top-down feedback occurs by comparing the current hypothesis against a normative (i.e., externally-driven) solution, and then either revising or discarding the current hypothesis and/or reframing the data. This reframing can occur by modifying the current frame. For instance, the model could change values (e.g., weights) associated with a given rule-like representation. An example would be increasing the likelihood of a given outcome based on the presence of a given feature. Reframing can also occur through utility learning, by reinforcing certain productions firing over others. For instance, penalizing productions that yield errors and reinforcing productions that test features which are diagnostic to the task. Reframing can apply to changing the hypothesis for the current data set as well as producing better hypotheses for future data sets.

While early evidence initially shapes the adoption of a frame, this frame can then shape how future evidence is recalled through the base-level and spreading activation mechanisms in the ACT-R architecture. In base-level activation, chunks that have been recalled in the past (which also spreads to related chunks through spreading activation) have higher activation, which make the recalling of similar data-chunks more likely in the future.

Before getting into more specifics regarding the functional and architectural analogues between the ACT-R cognitive architecture and sensemaking theory, we describe two tasks that instantiate the process.

3. The Tasks

The following tasks are designed to study the role of cognitive biases in sensemaking in the context of intelligence analysis. A facility identification task examined the ability of human participants to learn to analyze simulated geospatial images and correctly discriminate facilities in unlabeled images. Six group identification tasks tested the ability of human participants to correctly identify which group was responsible for an attack based on evidence from layers of data in geospatial images, and the application of probabilistic rules associated with the data interpretations.

The facility identification task was compared to both Bayesian normative solutions and human performance. The models of the group identification tasks were developed prior to gathering human data and serve as predictions of the kinds of biases humans may exhibit.

3.1 The Facility Identification Task

Participants were trained to identify four kinds of facilities in simulated geospatial images. Each image depicts a single facility (e.g., factory complex) composed of a set of discrete features (e.g., buildings). The three categories of features were: IMINT (image intelligence), representing buildings and other terrain such as roads and rivers; MASINT (measurement and signature intelligence), representing signals of radiation or chemical concentrations; and SIGINT (signals intelligence), representing communication transmissions.

The statistical breakdown of features was not even: there were nine unique IMINT features, seven that represented buildings, and two that represented water features. In contrast, there were only two kinds of MASINT features, while the SIGINT features were entirely homogeneous. In addition, each IMINT could appear at most one time in each image, whereas multiple instances of SIGINT and each MASINT could occur in each image. Additionally, each building (IMINT) could have attached to it zero or one piece of rooftop hardware. Each of the four facilities had different base rates for the occurrence of each of the possible features.

The experiment was divided into two phases: a training phase and a testing phase. In the training phase participants were presented with 48 annotated examples of each facility (192 total examples), 16 at a time (in a four-by-four grid). In the testing phase the participants were presented with single unlabeled images sequentially. For each image, participants were required to report a probability distribution over the four possible facilities indicating the likelihood that the image contained each of the facilities. For more details on the facility identification task and comparisons with human data, see Rutledge-Taylor et al., (2011; forthcoming).

3.2 The Group Identification Tasks

The group identification tasks were a series of six tasks in which the participants' were to predict which of four groups was responsible for an attack. Each task was presented spatially in a 100 x 100 grid (representing 30 square miles) on a computer screen. The critical feature in Tasks 1 to 3 was signals of activity (SIGACTs), which represented previous attacks by the groups. The pattern of SIGACTs for each group was defined by a group center of activity and a dispersion value. SIGACTs were produced probabilistically according to these definitions.

Task 1 consisted of 10 blocks of 10 trials. A trial consisted of a single SIGACT, represented as a group letter, appearing on the display. On the 10th trial of each block the group responsible was hidden, with the SIGACT represented as an empty square. Participants were required to assess the probabilities that each of the two groups was responsible for the attack, and then were asked to produce a forced choice response.

Task 2 consisted of 10 blocks of 20 trials, similar to those in task 1. The difference was that there were four groups instead of two. In addition, participants were not required to produce a forced choice response after giving their probability estimates. They were instead required to draw a circle for each group that defined the two-to-one ratio of the likelihood of an attack by the given group occurring inside versus outside their circle (e.g., their sphere of influence).

Tasks 3 to 6 (see Figure 2.1) added the complexity of calculating distance along road networks. In task 3, participants were still required to find group centers from a series of SIGACTs, but distance between SIGACTs was now to be judged "as the cow walks" along a road network. As such, tasks 3 to 6 involved visual problem solving (e.g., path planning and curve tracing; Lefevre, Dell'Acqua, Roelfsema, & Jolicoeur, 2011). Task 3 consisted of 10 blocks of 20 trials. Participants were required to produce a probability distribution for each group's likelihood of being responsible for the SIGACT on the last trial of each block.

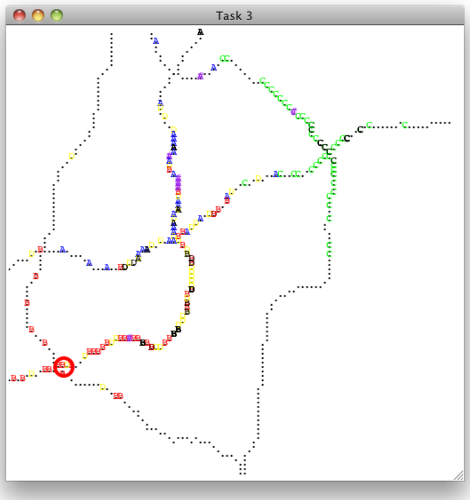

Figure 2.1. A sample screenshot of the group identification Task 3 model. The string of dots represents the road network, the letters represent individual SIGACTs, and the circle represents the model's focus of attention.

In tasks 4, 5 and 6, intelligence data was presented in layers: HUMINT gave the center of activity for each group (the participants were not required to judge this from individual SIGACTs in tasks 4 to 6); IMINT showed which of the roads in the network were the major roads and which were the minor roads; MOVINT showed which roads had dense traffic versus sparse traffic; SIGINT revealed information about whether a group was producing chatter or not; and SOCINT showed the territorial boundaries for the groups on the map.

Tasks 5 and 6 were very similar. In both tasks a road network with an anonymous SIGACT and the four group locations are presented. The participants' task is to update the probability distribution over the four possible groups responsible for the SIGACT after each layer is revealed. The basis for adjusting the probabilities is a set of rules that are provided to the participant. Each rule specifies changes in the relative likelihoods that the groups are responsible based on a piece of layer data. For example, if the SIGACT occurs on a major road, groups A and C are four times as likely to be responsible. In both tasks the HUMINT layer is provided first, and so the initial probability distribution is based on the relative distances between the group centers and the SIGACT. In task 5, the remaining layer are revealed, on at a time, in a random order. In task 6, the participant chooses, one at a time, three of the remaining four layers.

ACT-R models were produced for tasks 1, 2, 3, 5 and 6, prior to any human data being made available. Task 4 was not modeled as it was essentially a subset of task 5 and will be omitted from further discussion. As the models were generated prior to the collection of human data, they are predictive of human performance and provide an opportunity to examine and predict the influence of cognitive biases and provide possible solutions to reduce their impact on human judgments.

As of this submission, human data is still unavailable for the group identification tasks. The ACT-R model currently generates output probabilities for Tasks 2, 3, and 5 such that the highest group (of 4) is given a M = 49.8% probability (with the three other groups equally distributed, M = 16.7%). This is approximately 30% lower than a fully-rational Bayesian model (M = 81.2%) and reflects uncertainty due to conservatism and anchoring effects in the generation of group centers and probability judgments, and stochastic elements within the distance perception and path planning functions.

4. Sensemaking in the Identification Tasks

In the facility and group identification tasks, ACT-R modeled sensemaking processes at different levels of abstraction due to the increased task complexity in the group identification tasks. For the facility identification task, an ACT-R model learned which facility features were the most diagnostic to be attended to in order to correctly classify the facility. The model oscillates between the foraging and sensemaking loops as it acquires evidence, changes frames, and updates its facility identity hypothesis.

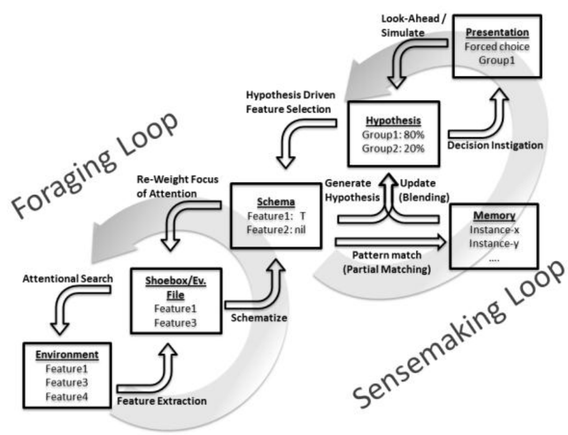

Figure 4.1. The ACT-R analogue to sensemaking.

As described in Figure 4.1, the foraging loop is the sequence of productions (represented as the arrows) which select new features from the environment (e.g., sensorimotor modules) and organize them in the imaginal buffer. In ACT-R, there is little distinction between the shoebox and evidence file, as similar productions determine which features are attended and harvested into declarative memory. The shoebox holds an initial selection of features extracted from the sensorimotor modules, which are then harvested into declarative memory via the imaginal module. Evidence files include a similar feature extraction from either the shoebox (memory) or environment (sensorimotor modules), but also have productions which may re-encode features from the shoebox and extract higher-order feature relations based on the currently-held frame/schema.

The sensemaking loop is the sequence of productions that retrieves a schema (e.g., a facility frame chunk) and generates a hypothesis for what facility is present (or group responsible) in a given task. In the facility identification task, the model oscillates between the two loops, seeking out more evidence and updating its hypothesis until a threshold for the expected utility (e.g., information gain) of making a decision is surpassed.

The facility identification task is an example of a case when a single chunk is sufficient for storing all relevant evidence in a frame. The frame was a single-chunk representation of the set of features to be attended to. Knowledge of the probabilities of the various features being present in a given instance was implicit in the set of chunks and their activations. The facility identification task is thus best described as a category-learning task as each frame can be interpreted as a category exemplar.

In the group identification tasks, the model encodes visual information in the foraging loop hierarchically. For instance, a chunk representing a road segment included a slot for road identification and two slots that each hold intersection-chunks, which in turn had slots for co-ordinate pair chunks containing *x* and *y* coordinates as slots. In addition, in tasks 5 and 6 there were several rule-like chunks (each corresponding to a layer) that specified how new evidence impacted the probability of a given hypothesis. Thus the current frame held by the model was represented by the current INT layer in the retrieval buffer, which would be used to update the current hypothesis (i.e., probability distribution).

4.1 Sensemaking in the Facility Identification Task

The ACT-R model of the facility identification tasks implements an adaptive foraging loop. The feature selection process is composed of two distinct phases. In the training phase the model studies a set of images for the purpose of learning which features are associated with each facility. In the learning phase the model must identify the facilities in images, and with feedback learn the optimal utilities for the various feature selector and decision instigation productions.

Learning frames was accomplished during the task's training phase. ACT-R modeled variants instantiating both rule-based and exemplar-based category learning. In this phase the model acquires examples of facilities and commits them to memory. In this case a frame is an abstraction of the accumulated exemplars that is realized during category assignment. The frame for a particular facility is thus the implicit knowledge that the model possesses about the probabilities of the various features being present in an instance. This is functionally similar to a rule with the implicit probabilities for each feature representing a set of conjunctions updating the likelihood of a given facility based on the presence or absence of a given feature.[1] The currently-held hypothesis is the probability that the current image depicts a given facility (and is provided to the model in the training phase).

Due to the feedback received during the training phase, the ACT-R model is constantly reframing by determining the utility of which features are the most diagnostic of the facilities. Feature selection is the process of deciding which features present in an image ought to be attended to, and which should not. Part of sensemaking theory is the ability to aggregate and distill information to maximize the availability of information within the context of working memory limitations. Feature selection, in part, addresses the issue of the working memory capacity for information.

It is presumed that the participants are unlikely to be able to attend to every available feature in every image due to various cognitive constraints. The normative probability that a feature should be selected is based on its utility in facility identification. The ACT-R model uses utility learning to develop implicit preferences for attending to some features over others. Utility learning provides rewards (or penalties) to productions based on their outcome. A positive reward is instigated after a facility is correctly identified, while a negative reward is instigated after an incorrect identification.

The productions of interest during utility learning are divided into two categories: feature selector productions and decision instigation productions. Each feature selector production is specific to a single IMINT feature and a specific intermediate hypothesis about what facility is represented in the given sector. This allows for the utilities of selecting features to be hypothesis-specific and represents the interplay between frames and hypotheses in the sensemaking loop.

Each decision instigation production is eligible to fire after a specific number of features have been selected. Once a decision instigation production has fired, a facility identification event occurs. If the identification is correct, the decision instigation production and all the feature selection productions that lead to the decision are rewarded. If the identification is incorrect, the same productions are penalized.

In the learning phase, the model alternates between updating the model's current hypothesis of what facility is present and selecting a new feature (or electing to stop encoding features). When selecting a feature, all of the

[1] In another paper (Rutledge-Taylor, Lebiere, Thomson, Staszewski & Anderson, forthcoming) we discuss similarities in performance between rule and exemplar-based models on this task.

feature selector productions that are eligible to fire compete. The production with the highest utility will fire, and the feature associated with the production will be added to the selected features stored in the imaginal buffer. This is analogous to acquiring additional data.

The decision to stop encoding features is governed by a set of decision instigation productions that compete against the feature selection productions. The decision instigation productions receive utility rewards and punishments, as do the feature selection productions (similar to the training phase). The model stops encoding features when the relevant decision instigation production fires instead of any of the eligible feature selection productions. This evolving production competition allows the model to learn the rational number of features to encode and represents the model's reframing based on whether the hypothesis was supported or contradicted.

When updating the current hypothesis (stored in the goal buffer), the facility is identified by recalling the facility frame from declarative memory that best matches the features selected so far. The value for the facility ID in the recalled frame is used to update the hypothesis maintained in the goal buffer. When the model stops encoding new features, the current hypothesis is output as the model's final categorization decision for the image. It is possible that, based on new features (reframing), the hypothesis may be updated (new probabilities) or rejected (by choosing a new facility type).

In summary, the process of feature selection in the facility identification task mirrors the sensemaking process. The foraging loop is analogous to the sequence of productions that selects new features from the available pool and organizes them in the imaginal buffer. The sensemaking loop is analogous to the sequence of productions that retrieves a facility frame chunk from DM and generates a hypothesis for what facility is present in the given image. The model oscillates between the two loops, seeking out more evidence and updating its hypothesis until the expected utility of making a decision is greater than that of collecting more data.

4.2 Sensemaking in the Group Identification Tasks

Unlike the facility identification task, the group identification tasks do not have a training phase and the rules for adjusting probabilities are explicitly provided to participants (and the model). The model thus has less opportunity for learning due to feedback (based on revising hypotheses). Instead, the group identification tasks require more general spatial judgments (such as path planning) and the application of multiple rules that do not fit the traditional definition of frame as a singular structure or representation in the sensemaking literature.

In the facility identification task it was also practical (and fit within the spirit of working memory limitations) to represent the relatively small set of features in a single chunk. In the group identification tasks it was neither practical nor cognitively plausible to represent the full set of spatial information within a single chunk in DM. Instead, chunks of spatial information are represented hierarchically. For instance, a group center (in task 3) is located on a road segment, which is made up of a location (co-ordinate pair) along a road segment. Road segments are defined by their endpoints (intersections) and general length and shape, which are also linked to locations.

The basic unit of evidence in tasks 1 to 3 is a SIGACT, which corresponds most closely with evidence files in the foraging loop. SIGACTs are perceived by the visual module, their location and group identity placed in the imaginal buffer, and at the end of each block, an estimate of each group center (and two-to-one boundary in task 2) is performed. This group center and boundary estimate is a kind of spatial frame (insofar as it predominantly contains spatial information). In the model, the group center is calculated for each group in a separate chunk, thus the current frame of the model incorporates four chunks (one for each group center). Using these spatial frames, a hypothesis (i.e., the set of probabilities) is generated and compared against the provided feedback. The model has only a limited ability to reframe because it only receives feedback (i.e., ground truth, not probability distribution) at the end of each block. As such, reframing occurs when the model updates the group center estimate (in the subsequent block) with the identity and location of the target SIGACT from the previous block.

In tasks 3 to 6, distance (for the purpose of generating group centers from SIGACTs and between each group center and a target SIGACT) was not calculated using a 'crow-flies' estimate, but instead by estimating the length of the path along the road network. Due to the complexity of the road network, more than one path could be chosen. As such, the path-planning processes within the model could be seen as their own self-contained sensemaking process (implemented as a non-deterministic hill-climber). Foraging involved the perception of the possible paths at each intersection, framing involved the storage of the path, hypothesis testing involved mentally traversing a candidate road segment, and reframing occurred when the model needed to backtrack to a previous segment (due to hitting a dead-end or detecting that it had gone in a loop). The hypothesis included the currently-held distance estimate, and was revised when a new candidate road intersection was added to the path.

Tasks 5 and 6 use a more general model of hypothesis testing and reframing due to application of multiple layers of INTs. The spatial frames (from task 3) mapping group centers to a target SIGACT now represent a single layer (the HUMINT layer). There is little foraging to be done because the rules and group centers are provided as input to the model. Sensemaking is even more prevalent in task 6 because participants are able to choose their next INT layer (in task 5, three layers are provided in a random order). The model uses utility learning (similar to facility identification) to reward the model when it chooses a layer that leads to a correct probability distribution.

A frame is generated when the first layer of information (HUMINT) is applied to existing group centers. Using this frame, an initial hypothesis is generated (i.e., the initial probabilities for each group). Reframing occurs when an additional layer (SOCINT, SIGINT, MOVINT, or IMINT) is applied, which then revises the current hypothesis. Acquiring additional information occurs externally to the participant (and model) in task 5 because the layers are presented randomly; however the participant (and model) may reframe based on this additional evidence. In task 6, however, the model reframes by choosing a layer based on the current hypothesis. For instance, the model might select SOCINT when the probability distribution is flat, MOVINT or IMINT when trying to dissociate two alternatives, and SIGINT when the distribution is steep. The current hypothesis is revised when the next INT layer is applied (i.e., when the rule is applied which in turn revises the probabilities), and the layers that lead to correct classifications are rewarded.

In summary, the closest equivalent to a single frame in the group identification tasks is an organized representation of the input (i.e., the group centers in Tasks 1 to 3; the HUMINT, SOCINT, SIGINT, MOVINT and IMINT layers in Tasks 4 to 6). This definition of a frame preserves the meaning of the sensemaking processes described in the facility identification task.

More specifically, learning a frame corresponds to the accumulation of evidence supporting the hypothesis of a group being responsible for a SIGACT. This involves the accumulation of group centers estimates (in tasks 1 to 3), the dispersion of attacks (task 2), and generating initial probabilities based on rules (Tasks 4 to 6). Importantly, the model assumes that some spatial and mathematical mapping chunks already exist in memory, and reflect general experience/competencies.

Generating a hypothesis corresponds to the initial probability distribution assigned to each group. The probabilities are an evaluation of how probable it is that a particular group was responsible for a target SIGACT, based on how well the features of the SIGACT (e.g., location) match the characteristics of each group's frame. Reframing occurs whenever feedback occurs and when a new INT layer is applied.

4.3 Modeling Biases in Sensemaking

In sensemaking, biases are usually identified as heuristics in data gathering within the foraging loop and in frame (re-)encoding in the sensemaking loop (e.g., availability heuristic; Klein, Moon, & Hoffman, 2006). Under this interpretation, biases result mainly from architectural constraints (e.g., working memory, attention). Biases, however, may also occur in the sensemaking process due to the nature of the task and how the limitations of human memory and attention degrade performance in systematic ways. When modeling human performance, how the computational model is constructed will influence how biases are reflected in the model. For instance, the bias may be an emergent property of the architecture, or it may be due to the specific strategies of the model itself.

Thus, in cognitive modeling biases may arise from a combination of task demands, architectural limitations (e.g., one item per buffer in ACT-R), and modeler preferences. For instance, confirmation bias (an overly certain belief in the leading hypothesis) may be due to an explicit strategy coded by the model to always focus on the group with the highest probability (e.g., choose the SIGINT rule for that group) or it may be implicit in the architecture of the model (e.g., spreading activation from contextual cues increasing the probability of some features being recalled). This was a factor in the feature selection model of the facility identification task. By maintaining a chunk representing the current facility hypothesis in a slot of the goal buffer, the model was biased towards maintaining that hypothesis because exemplars in DM of the same facility would receive a boost in activation via spreading activation and thus make a greater contribution to updating the current hypothesis. It may also be possible to equate model-level biases with biases resulting from explicit human strategies (Gigerenzer & Gaissmaier, 2011).

Base-rate neglect is another example of a bias which can be a result from either the task, implicit (i.e., architectural), or strategic levels. At the task level, higher levels of complexity can cause base-rate neglect due to the sheer volume of stimuli to store and subsequently recall. To reduce both memory load and processing time, some features need to be abstracted. Similarly, at the implicit level, working memory is generally seen to have a capacity of 7 +/- 2 chunks of information available (Miller, 1956). Once this memory capacity is exceeded, some information needs to be either abstracted or discarded. Finally, at the strategic level, there are coding choices that may be made based on an analysis of task difficulty and an awareness of architectural constraints. For instance, when grouping many dots presented sequentially on a display it may be an explicit strategy to ignore base-rate and focus on grouping elements together.

The goal of cognitive modeling should primarily be to account for biases due to the interplay between the task and architectural levels. Those biases should be considered emergent properties of the model. Explicit strategies would be represented in the model as design choices in terms of the specific productions and chunk types utilized. Explicit reasoning strategies may influence performance (and often reflect architectural limitations), but should not be relied upon as the mechanism to instantiate biases (especially biases considered to be due to automatic processes).

5. Discussion

The advantage of standardizing sensemaking processes with ACT-R is that it provides a common framework for model comparison in extant sensemaking tasks and possible model re-use between tasks. A limitation of cognitive modeling is that models can rarely be generalized or re-used to adapt to a different task. By standardizing the elements of sensemaking with ACT-R, we are providing a roadmap to generate a general cognitive model of sensemaking, capable of making predictions regarding human performance (as opposed to fitting to extant human data). A generalized model of sensemaking is under development, which is based on the principles described in this document.

In summary, the present paper has provided a possible framework and some suggestions for how to encode stimuli and represent hypotheses in ACT-R that would instantiate sensemaking processes. While it was beyond the purview of this discussion to provide a more in-depth link between the individual sensemaking elements and the ACT-R architecture, it does provide a framework for a general cognitive model of sensemaking.

6. Acknowledgement

This work is supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract number D10PC20021. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwith-standing any copyright annotation thereon. The views and conclusions contained hereon are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government.

7. References

Anderson, J. R., & Betz, J. (2001). A hybrid model of categorization. *Psychonomic Bulletin & Review, 8*, 629-647.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y. (2004) An integrated theory of Mind. *Psychological Review, 111*, 1036-1060.

Anderson, J. R., and Lebiere, C. (1998), *The atomic components of thought*, Erlbaum, Mahwah, NJ, 1998.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology,* 62, 451–482.

Kahneman, D., & Treisman, A. (1984). Changing views of attention and automaticity. In R. Parasuraman, D.R. Davies & J. Beatty (Eds.), *Variants of attention* (pp. 29-61). New York: Academic Press.

Klein, G., Phillips, J. K., Rall, E., & Peluso, D. A. (2006). A data/frame theory of sensemaking. In R. R. Hoffman (Ed.), *Expertise out of context: Proceedings of the 6th International Conference on Naturalistic Decision Making*. Mahwah, NJ: Lawrence Erlbaum & Associates.

Klein, G., Moon, B., & Hoffman, R. (2006). Making Sense of Sensemaking 1: Alternative Perspectives. *Intelligent Systems*, 21 (4), 71.

Lefebvre, C., Dell'Acqua, R., Roelfsema, P., & Jolicœur, P. (2011). Surfing the attentional waves during visual curve tracing: Evidence from the sustained posterior negativity. *Psychophysiology*, 48, 1-7.

Marr, D. (1982), *Vision: A Computational Approach*, San Francisco, Freeman.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review*, *63*, 81-97.

Pirolli, P. & Card, S. K. (2005). The sensemaking process and leverage points for analyst technology. In the *Proceedings of the 2005 International Conference on Intelligence Analysis*. McLean, VA: Office of the Assistant Director of Central Intelligence for Analysis and Production.

Pirolli, P., Fu, W.T., Reeder, R. and Card, S. K. (2002). A User-Tracing Architecture for Modeling Interaction with the World Wide Web. *Advanced Visual Interfaces*, Trento, Italy.

Rutledge-Taylor, M. F, Lebiere, C., Vinokurov, Y., Stazewski, J., & Anderson, J. R. (2011). Bridging the gap: A neurally plausible functional model of sensemaking. In Samsonovich, A. V., Johannsdottir, K. R., Chella, A., & Goertzel, B. (Eds.). *Proceedings of the Second Annual Meeting of the BICA Society.* Fairfax: IOC Press.

Rutledge-Taylor, M. F, Lebiere, C., Thomson, R., Stazewski, J., & Anderson, J. R. (forthcominng). A Comparison of Rule-Based versus Exemplar-Based Categorization Using the ACT-R Architecture. *The 21st Behavior Representation in Modeling & Simulation Conference.*

Author Biographies

ROBERT THOMSON is a Post-Doctoral Researcher at Carnegie Mellon University in Pittsburgh, PA. His main research interests are distance perception and spatial organization, serial memory, and biases in reasoning.

CHRISTIAN LEBIERE is research faculty in the Psychology Department at Carnegie Mellon University. His main research interests are computational cognitive architectures and their applications to psychology, artificial intelligence, human-computer interaction, decision-making, intelligent agents, cognitive robotics and neuromorphic engineering.

MATTHEW RUTLEDGE-TAYLOR is a Post-Doctoral Researcher at Carnegie Mellon University in Pittsburgh, PA.

JAMES STASZEWSKI is a Professor of Psychology at Carnegie Mellon University in Pittsburgh, PA.

JOHN R. ANDERSON is a Professor of Psychology at Carnegie Mellon University in Pittsburgh, PA.