SAICMP94.0010
Copy 1 of 2

Application–Oriented Receiver Certification (U)

June 9, 1994

Science Applications International Corporation
An Employee-Owned Company

Presented to:

SG1J [redacted]

Defense Intelligence Agency
P.O. Box 440
Odenton, MD 21113

Contract MDA908–93–C–0004

Submitted by:

Edwin C. May, Ph.D.
Science Applications International Corporation
Cognitive Sciences Laboratory
P.O. Box 1412
Menlo Park, CA 94025

ABSTRACT (U)

(S/NF) We describe a three-tier procedure to certify the skill and ability of operational-oriented practitioners of anomalous cognition. The first tier is the most relevant to operations. In it, we suggest a 5-level qualitative assessment criteria, which is based upon ground truth supplied by the customer. In addition, we urge that all operational tasks be divided into appropriate categories of tactical and strategic intelligence. Thus, the 5-level criteria will be applied within a given category, and will, therefore, be mission sensitive. We describe this method in detail and suggest minimum and reasonable certification criteria for this tier. If a practitioner fails this first certification, then we suggest a second tier, which is also operationally relevant. That is, the practitioner provides data in what he or she believes is a true operational problem. However, simulated operational targets, in which complete ground truth is available, are used in this tier. We provide a detailed quantitative and analytical method of evaluating performance in what is called a test-bed environment. As in the first tier, we suggest certification minimums. Finally, if the practitioner fails the first two tiers, we suggest a laboratory experiment as the final attempt for certification. We present the details of the laboratory techniques and provide certification minimums and rationals. If a given practitioner cannot be certified by the recommended three-tier method, we suggest that he or she be dismissed from the operational unit.*

* (S/NF) This report constitutes the deliverable for the Operational Certification Task under contract MDA908-93-C-0004.

I. INTRODUCTION (U)

(S/NF) Anomalous cognition (AC) is defined as the acquisition, by mental means alone, of information that is otherwise secured by distance, time, or shielding. The existence of AC has been established by research in mainstream open literature (Puthoff and Targ, 1976 and Bem and Honorton, 1994) and in the classified literature in over 150 reports (May and Luke, 1991). Attempts to use AC against operationally sensitive problems of National Security interest began in 1972 with a contract with the Central Intelligence Agency (CIA) and continues to date under the auspices of the Defence Intelligence Agency (DIA).

(S/NF) We have often recommended that operational receivers* not be chosen from unit personnel. There is a long history of research which indicates that performance anxiety, boredom, or psychological "burn out" are contributing factors to a steady, but significant, decline of performance. In addition, we find that receivers are less willing to "risk" their impressions which may eventually contribute to the disruption of unit cohesiveness. Regardless of the receivers' location, it is paramount to subject their output to continuing performance review. Such a review, or certification, can guide the use of receiver resources effectively and determine if a given receiver should remain with the program. We have required a preset minimum level of performance for our research receivers for the last 10 years.

(S/NF) In developing an operationally relevant certification procedure, we must consider a closely associated concept; the intelligence utility of AC-derived information. The assessment of intelligence is, in itself, problematical, and one approach, which is based on sophisticated optimization strategies, has made significant progress toward that end (Taguchi and Phadke, 1984; Phadke and Dehnad, 1987; Taguchi, 1993). It is beyond the scope of this report to provide a description and analysis of what is known as the Taguchi method, but we include it here for completeness. Rather, we will assume that some valid intelligence assessment tool exists and focus our attention on the problem of receiver certification, instead.

* (S/NF) We use the term *receiver* to indicate source, subject, or participant in AC operations.

II. METHODS OF CERTIFICATION (U)

(S/NF) In this discussion, we use a top-down approach; that is, starting with the intelligence product we evolve toward an exclusively laboratory certification.

1. Certification by Example (U)

(S/NF) Perhaps the only valid measure of receiver certification for operational AC is a satisfied customer. One advantage of certification by example is that a valid, independent intelligence utilization measure (e.g., Taguchi method) is not required. Each customer independently defines whether or not the AC data was useful. Still, a number of requirements must be fulfilled before such a certification procedure can be implemented, and the procedure for which should be task sensitive. That is, one receiver might be certified for some operational categories but not for others.

1.1 Scoring Procedure (U)

(S/NF) Broad categories of AC-intelligence must be identified. They should be dynamic (i.e., as requirements change, topics are added to or dropped from the list) and should be divided into tactical and strategic items. Although there is not a sharp boundary between these two, tactical intelligence problems tend to be more time critical than strategic ones. For example, location of individuals within a small period of time, or the identification of major events (e.g., missile firing, terrorists' attack) might be included among the tactical intelligence categories; while facility floor plans, facility purpose, or nuclear production schedules are more appropriate for strategic categories.

(S/NF) Once a reasonable set of categories have been identified, an in-house quality assessment based upon feedback (i.e., ground truth) supplied by the customer must be developed. We emphasize that this assessment is made at the total task level rather than on an item-by-item basis. This last point is very important. An excellent example of AC may score well item-by-item; however, for a variety of reasons, the data might not be of any intelligence value. For example, an AC-derived floor plan, which may be accurate to the nearest centimeter, is of no strategic value because the floor plan may be obtained by HUMINT sources and, thus the AC data provides no new or particularly confirming information. On the other hand, AC data that would not meet laboratory criteria for excellent performance, might provide a single element that serves as a tip-off and cracks a particularly intractable intelligence problem. In both cases, an item-by-item analysis will not reflect the intelligence utility of the data.

(S/NF) Suppose we invent a 5-level task assessment scheme as shown in Figure 1. (Basic research has shown that humans are not capable of reliably separating seven ± two elements in subjective assessment tasks (Dawes, 1988), thus we have chosen five levels for our intelligence utility scale.)

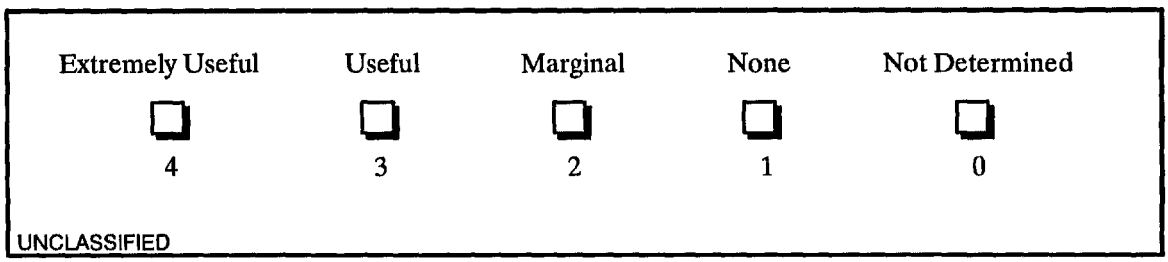

Figure 1. (S/NF) Intelligence Utility Scale for AC-Data

We emphasize that this scale is to be used by an in-house analyst—not a customer analyst. For each intelligence task where ground truth can be obtained per each receiver, an analyst must assign a value based upon a subjective assessment of the customer report and the ground truth. Ideally, the same analyst would make such assessments for all receivers in the unit.

(S/NF) Over time for a given receiver, an on-line database can keep track of the percentage of tasks that received each of the possible utility scores. Figure 2 shows an example of two intelligence utility records for a specific tactical intelligence category (e.,g., event recognition) for receivers α and β.

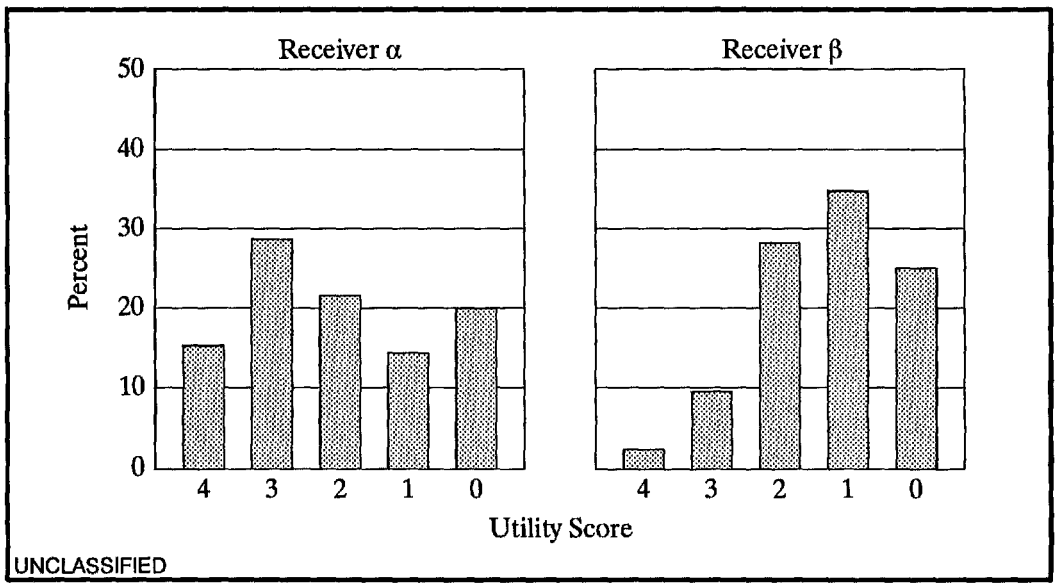

Figure 2. (S/NF) Utility Record on a Tactical Intelligence Category for Receivers α and β.

(U) The total percentage must sum to 100 for each receiver's record. By visual inspection, receiver α is much better, in the long term, for this particular category. A more sensitive figure for overall performance is the numerical average of these utility scores excluding zero. That is, of all the operations where ground truth was available, what is the performance level? In our example the averages are 2.561 and 1.728 for receiver α and β, respectively.

1.2 Certification Matrix (U)

(S/NF) Table 1 shows a sample certification matrix. This matrix contains one row for each receiver and one column for each intelligence category, which we have described above. The value for a given receiver and a given category is the average of the in-house assessment as indicated in Figure 2.

Table 1

Certification Matrix (U)

Receiver	Category				
	A	B	C	D	E
α	2.351	1.433	1.212	1.843	1.095
β	1.222	1.629	1.404	1.057	1.015
δ	1.193	2.531	1.094	1.706	1.741
ε	1.871	2.298	1.984	1.767	1.151

UNCLASSIFIED

(S/NF) Suppose we set a liberal threshold for certification of *1.75*. That is, over many operational AC sessions, a receiver must produce data that is on the average deemed to be close to marginally useful.* We suggest that as many sessions as possible be included in the average so that an accurate assessment can be made. The underlined values indicate those that exceed this *1.75* threshold. With this criteria, receiver α passes for categories *A* and *D* but fails in the others. Similarly, receiver δ provides useful information in category *B*, and receiver ε is good in all categories except *E*. We notice that no receiver performs in category *E*, which indicates that either this category should be dropped and such operational tasking should be rejected, or that a search should be initiated to find a receiver who may be proficient in this category.

(U) Another useful concept emerges from this matrix. The indicated proficiencies can guide the project manager to assign receivers only to tasks in which they have a demonstrated proficiency. Thus, overall production will improve.

(S/NF) Finally, we notice that receiver β has failed the certification for all current intelligence categories. While it may be tempting to dismiss receiver β, our top-down certification procedure suggests a different approach. It is possible that this receiver may be proficient on some other category not currently being considered.

(S/NF) The next level of certification involves simulated operations (i.e., test-bed experiments) in which total ground truth is known, but the receiver is unaware of the "test" nature of the activity.

* (U) A more conservative and demanding threshold might be 2.25.

2. Test-bed Certification (U)

(S/NF) We have been conducting operational simulation experiments for a number of years (May, 1988; May 1989). These test-bed experiments differ from true operations in that total ground truth is known in advance. Other than that, the AC sessions are conducted as if the session were an actual intelligence operation. The candidate receiver can use the methods he or she finds comfortable and the targeting techniques that are generally used in operations can be maintained. Although it is not a requirement, better results can be obtained if the candidate receiver is unaware that the session is a test-bed certification trial.

(S/NF) Since the test-bed target is known in its entirety, a list of items can be constructed that would be of intelligence interest. We illustrate this approach to receiver certification with one of our test-bed experiments.* We constructed three categories of items: (1) Functions of the Site, (2) Physical Relationships, and (3) Objects. Table 2 shows a partial list of these three types of items for our test-bed experiment in which the target system was a 50 MeV, 10^4 ampere electron beam being projected into air (May, 1988). The complete list spans many pages.

Table 2.

Partial Element List for a Test-bed Experiment (U)

Target/Response Element	w	T(μ)	R(μ)
Functions (1.0)			
Directed Energy	5	1.0	0.9
Test Experiment	2	1.0	1.0
Noise Generation	1	0.4	0.6
Operation in Space	1	0.0	1.0
Relationships (0.75)			
Power Source Above Beam Line	1	1.0	0.0
Electrons Flow Through Beam Line	1	1.0	0.7
Pipes in and out of Sphere	1	0	1.0
Objects (0.5)			
External Electron Beam	2.5	1.0	0.0
High Security Area	1	1.0	1.0
Bundled Metal Rods	1	0.0	1.0

SECRET/NOFORN

(S/NF) To provide an accurate certification measure, two types of data must be incorporated into such a list; an *a priori* list of items that are definitely part of the target and items that are mentioned by the receiver that were not recognized as being part of the target. In Table 2, we have indicated overall weighting factors of 1.0, 0.75, and 0.5 for functions, relationships, and objects, respectively. Meaning that, in this experiment, the client was primarily interested in functions. Depending upon the task, the formalism will accept any appropriate weighting factors. The column *w* is a within-group weighting fac-

* (U) Of course, in implementing this part of the certification procedure, the project director would construct a different list, which is mission and target dependent.

tor. The item *Directed Energy* is five time more important than is *Noise Generation*. $T(\mu)$, the target score, represents the degree to which the item is present in the target. For example, although *Noise Generation* is present in the target, it is roughly 40% apparent; whereas *Pipes in and out of Sphere* is not present at all. $R(\mu)$, the response score, is the degree to which the analyst is convinced that the element is indicated in the response. For example, the analyst was 90% convinced that the receiver meant *Directed Energy* even though it was not specifically mentioned. All items that are specifically mentioned receive an $R(\mu) = 1$. Notice that we include all items mentioned by the receiver regardless if the item was present in the target. We set their relative weights all equal to one.

(U) To arrive at a meaningful number from these data, we use fuzzy set formalism (May, Utts, Humphrey, Luke, Frivold, and Trask, 1990). We compute the accuracy and the reliability of the response to the target system. The accuracy is the fraction of items in the target that were described correctly, and the reliability is the fraction of items in the response that were present in the target system. It is possible to obtain a very accurate description with poor reliability. Suppose the receiver mentioned everything that can be found in an encyclopedia as his or her response. In principle, nearly all aspects of the target might be mentioned; however, a large number of response items would not be present in the target. The certification number must be related to the accuracy and reliability. Formally, the accuracy and reliability are defined by:

$$Accuracy = \frac{\sum_{j=1}^{N} W_j \, Min\left[T_j(\mu), R_j(\mu)\right]}{\sum_{j=1}^{N} W_j \, T_j(\mu)},$$

$$Reliability = \frac{\sum_{j=1}^{N} W_j \, Min\left[T_j(\mu), R_j(\mu)\right]}{\sum_{j=1}^{N} W_j \, R_j(\mu)}, \tag{1}$$

where N is the total number of elements in the evaluation form; T_j and R_j are the target and response score for element j; and W_j is the product of the within-group weight, w, and the group weight. For example, in the *Functions* group the w are equal to the W because the functions weight is one. Since the *Relationships* group weight is *0.75*, the within-group weights shown in Table 2 must all be multiplied by *0.75* to form the W_j for those elements in this group.

(U) To be sensitive to the interplay between *Accuracy* and *Reliability*, we propose that *Certification* = *Accuracy* × *Reliability*.

(U) To illustrate the use of Equations 1, we demonstrate how to compute the *Accuracy* from the data in Table 2. We note that *Min* function means to select the smaller of the target and response score. There are 10 items in Table 2, so the *Accuracy* = *[1* × *(5* × *0.9* + *2* × *1* + *1* × *0.4* + *1* × *0)* + *0.75* × *(1* × *0* + *1* × *0.7* + *1* × *0)* + *0.5* × *(2.5*× *0* + *1* × *1* + *1* × *0)* *]* divide by *[1* × *(5* × *1* + *2* × *1* + *1* × *0.4* + *1* × *0)* + *0.75* × *(1* × *1* + *1* × *1* + *1* × *0)* + *0.5* × *(2.5*× *1* + *1* × *1* + *1* × *0)* *]* = *7.925 / 10.65* = *0.744*. Similarly, we compute the *Reliability* = *0.764*, and *Certification* = *0.568*. In our test-bed experiment, that *Accuracy*, *Reliability*, and *Certification* were *0.81*, *0.76*, and *0.61*, respectively.

(U) Random utterances compared to random targets roughly yield *0.3* for both *Accuracy* and *Reliability*. That is, approximately 1/3 of whatever is said can be found in any target and 1/3 of any target can be described regardless of what is said. An approximate *Certification* of *0.1* would represent chance matches.

(S/NF) For this second-level, the test-bed certification procedure, we suggest a *Certification* value of three times chance, or *0.3*, be the absolute minimum that would allow an operational receiver to remain as a resource. If the receiver's score is routinely less than *0.3* in a series of test-bed trials, we suggest a laboratory experiment for the final attempt at certification before the receiver is dismissed from the unit.

3. Laboratory Certification (U)

(S/NF) We propose that laboratory certification be the "court of last resort" for an operational receiver. Although it is sometimes argued that operational AC is fundamentally different than laboratory AC, the experience and research spanning 20 years in our laboratory is unable to confirm this idea. In fact, our best receivers perform equally well in laboratory experiments and operations. This conclusion is drawn from many hundreds of operational trials conducted during this time.

(U) One advantage of a laboratory certification procedure is that the protocols and assessment techniques are well understood. Many different laboratories have validated a variety of techniques during the last 20 years (Honorton and Harper, 1974; Jahn, 1982; May, Utts, Humphrey, Luke, Frivold, and Trask, 1990; Lantz, Luke, and May, 1994).

(U) For a laboratory certification to be valid, it must incorporate the current research understanding as much as possible. With this in mind, we suggest that a candidate receiver participate in 24 laboratory trials, which are conducted at a rate of no more than three per week. The complete protocol for a single trial is as follows:

(1) The receiver and a monitor (i.e., a skilled interviewer) enter a quiet and isolated room.

(2) An assistant randomly selects one target from a pre-defined set. For these targets, we suggest 100 photographs from the *National Geographic* magazine of natural and man-made scenes. These photographs should be divided into 20 packets of 5 targets each such that within a packet, the photographs are as different from one another as possible. Please see May et al. (1990) for a complete description of a target pool construction technique.

(3) At a pre-arranged time, the receiver, who is unaware of the selection, records his or her impressions of the target with written words and drawings. The monitor, who must also be "blind" to the target selection, is free to guide the receiver. In particular, the monitor is to keep the receiver from analyzing the impressions whenever possible.

(4) After the AC data is complete, the monitor copies the response, secures the original, and obtains the target photograph for feedback. During the feedback time, the monitor and receiver completely debrief the experience, and identify correspondence between the response and target.

(U) At the end of 24 such trials, the records include 24 responses, target pack numbers, and within-pack target numbers. A trained analyst, who has no prior knowledge of any of the data, must conduct the certification analysis. He or she will know the target pack from which the intended target for each trial was selected. The procedure for the analysis of each response is as follows:

(1) Regardless of the quality of the given response, the analyst must subjectively decide which of the five targets within the pack best matches the response.

(2) Having chosen the target for the best match, the analyst next chooses the target which is the second best match.

(U) The analyst continues in this way until the 5th best target matche has been determined. The position of the intended target is called the rank. That is, if the analyst believed that the intended target was the second best match, a rank of two is assigned for that trial. At the end of 24 trials, the analyst has produced 24 rank numbers. Adding these together and dividing by 24 produces the average rank. The effect size (i.e., certification value) is given by:

$$ES = \frac{(3 - \textit{average rank})}{\sqrt{2}}. \qquad (2)$$

(S/NF) The band of effect sizes in which there is a 95% confidence that the true value resides is $ES \pm 0.336$. We suggest, therefore, that a minimum value for a valid certification effect size should be 0.4, and a more reasonable one, which indicates excellent AC performance for operations, should be 0.6. Our best receivers produce effect sizes of 0.7.

(S/NF) If a candidate receiver fails to reach even the minimum effect size, we recommend that he or she be barred from participating in operational tasks.

III. CONCLUSIONS (U)

(S/NF) We have described a three-level certification procedure for operational-oriented receivers. We believe the suggested methods are sensitive to each receivers individual techniques, yet provide quantitative evaluations that have been approved by our panel of scientific experts (i.e., the Scientific Oversight Committee). While it is our firm conviction that no personnel should be assigned as dedicated receivers, our recommended certification technique provides objective criteria for their continuation in that capacity.

REFERENCES (U)

Bem, D. J. and Honorton, C. (1994). Does psi exist? Replicable evidence for an anomalous process of information transfer. *Psychological Bulletin.* **115**, No. 1, 4-18, UNCLASSIFIED.

Dawes, R. M. (1988). *Rational choice in an uncertain world.* Harcourt Brace Jovanovich, New York, NY, UNCLASSIFIED.

Honorton, C. and Harper, S. (1974). Psi-mediated imagery and ideation in an experimental procedure for regulating perceptual input. *The Journal of the American Society for Psychical Research.* **68,** 156-168.

Jahn, R. G. (1982). The persistent paradox of psychic phenomena: an engineering perspecitve. *Proceedings of the IEEE.* **70**, No. 2, 136-170.

Lantz, N. D., Luke, W. L. W., and May, E. C. (1994). Target and sender dependencies in anomalous cognition experiments. Submitted for publication in the *Journal of Parapsychology.*

May, E. C. (1988). An application oriented remote viewing experiment (U). Final Report, SRI International, Menlo Park, CA, SECRET/NOFORN.

May, E. C. (1989). An application oriented remote viewing experiment (U). Final Report, SRI International, Menlo Park, CA, SECRET/NOFORN.

May, E. C., Utts, J. M., Humphrey, B. S., Luke, W. L. W., Frivold, T. J., and Trask, V. V. (1990). Advances in remote-viewing analysis. *Journal of Parapsychology,* **54**, 193-228, UNCLASSIFIED.

May, E. C. and Luke, W. L. W. (1991). A proposal for research of anomalous mental phenomena. Submitted to DIA.

Phadke, M. S. and Dehnad K. (1987). Optimization of product and process design for quality and cost. *Quality and Reliability International*, **4**, 103-112, UNCLASSIFIED.

Puthoff, H. E. and Targ, R. (1976). A perceptual channel for information transfer over likometer distances: Historical perspective and recent research. *Proceedings of the IEEE*, **64**, No. 3, 329-354, UNCLASSIFIED.

Taguchi G. and Phadke, M. S. (1984). Quality engineering through design optimization. *Conference Record,* GLOBECOM84 Meeting, IEEE Communications Society, Atlanta GA 1106-1113, UNCLASSIFIED.

Taguchi G. (1993). *Taguchi Methods*, Quality Engineering Series, **4**, UNCLASSIFIED.