*Journal of Parapsychology*, Vol. 54, September 1990

# ADVANCES IN REMOTE-VIEWING ANALYSIS

BY EDWIN C. MAY, JESSICA M. UTTS, BEVERLY S. HUMPHREY, WANDA L. W. LUKE, THANE J. FRIVOLD, AND VIRGINIA V. TRASK

**ABSTRACT: Fuzzy set technology is applied to the ongoing research question of how to automate the analysis of remote-viewing data. Fuzzy sets were invented to describe, in a formal way, the subjectivity inherent in human reasoning. Applied to remote-viewing analysis, the technique involves a quantitative encoding of target and response material and provides a formal comparison. In this progress report, the accuracy of a response is defined as the percent of the intended target material that is described correctly. The reliability is defined as the percent of the response that was correct. The assessment of the remote-viewing quality is defined as the product of accuracy and reliability, called the figure of merit. The procedure is applied to a test set of six remote-viewing trials. A comparison of the figures of merit with the subjective assessments of 37 independent analysts shows good agreement. The fuzzy set technology is also used to provide a quantitative definition of target orthogonality.**

Human analysts are commonly used to evaluate free-response data. Although there are many variations, the basic idea is that an analyst, who is blind to the actual result, is presented with a response and a number of target possibilities, one of which is the intended target. The analyst's task is to decide what is the best response/target match, and frequently includes rank-ordering the targets from best to worst correspondence with the response. It is beyond the scope of this report to provide a critical review of the extensive literature on this topic.

One aspect, however, of this type of evaluation is that analysts are required to make global judgments about the overall match between a complex target (e.g., a photograph of a natural scene) and an equally complex response (e.g., written words and drawings). In a recent book, Dawes (1988) has discussed various decision algorithms in general and the difficulty with global techniques, such as those used in rank-order evaluation, in particular.[1] According to Dawes, the research results suggest that global decisions of this type are not as good as those based on smaller subelements that are later

[1] We are indebted to Professor D. Bem, Cornell University, for directing us to this valuable source of information.

combined. (See Dawes, 1988, chap. 10, for references to the research.) Humans appear to be capable of deciding what the appropriate variables should be in complex decision processes, but they have proved to be unreliable at combining these variables to arrive at a single decision. Linear algorithms are consistently better at this latter task. Therefore, it seems prudent to develop evaluation techniques that are less sensitive to global decision processes and rely on combinations of more restrictive decisions.

Honorton (1975) has pointed out an additional difficulty inherent in a global rank-order approach. Asking an analyst to rank-order a small set of target possibilities converts the free-response experiment into a forced-choice one, at least on the part of the analyst. It is obvious that in doing so, much quantitative information is lost. For example, a near perfect correspondence between response and target will receive only as much "credit" as one that just barely allowed an analyst to discriminate among the possibilities.

If multiple analysts are used, addition problems arise concerning interanalyst reliability. If an individual analyst judges a number of responses in a series, within-analyst consistency becomes an individual problem.

To address these difficulties, various computer-automated procedures have been suggested in an attempt to reduce the interanalyst reliability while increasing within-analyst consistency. For examples, see Honorton (1975), Humphrey, May, Trask, and Thomson (1986), Humphrey, May, and Utts (1988), Jahn, Dunne, and Jahn (1980), May (1983), May, Humphrey, and Mathews (1985), and Targ, Puthoff, and May (1977).

In this paper we present the current status of an ongoing research topic. We are not yet ready to propose that the techniques described here be used for free-response analysis; however, we hope to inspire the community to develop a proper set of subvariables so that the problems inherent in global decision processes can be avoided.

Finally, we present a successful application of the mathematical techniques for quantifying target orthogonality for a complex target pool.

*Background*

Substantial progress has been made in methods for evaluating remote-viewing experiments since the publication of the initial remote-viewing (RV) effort at SRI International (Puthoff & Targ,

1976). This paper outlines some of the progress and presents the details for one particular method.[2]

Two basic questions are inherent in the analysis of any remote-viewing data, namely, how is the target defined, and how is the response defined.

In a typical outbound RV experiment, definitions of *target* and *response* are particularly difficult to achieve. The protocol for such an experiment dictates that an experimenter travel to some randomly chosen location at a prearranged time; a viewer's task is to describe that location. One method of trying to assess the quality of the RV descriptions in a series of trials is to require that an analyst visit each of the sites and attempt to match responses to them. While standing at a site, the analyst has to determine not only the bounds of the site, but also the site details that are to be included in the analysis. For example, if the target location was the Golden Gate Bridge, the analyst would have to determine whether the buildings of downtown San Francisco, which are clearly and prominently visible from the bridge, were to be considered part of the target. The RV response to the Golden Gate Bridge target could be equally troublesome, because responses of this sort are typically 15 pages of dream-like free associations. A reasonable description of the bridge might be contained in the response; it might be obfuscated, however, by a large amount of unrelated material. How is an analyst to approach this problem of response definition?

The first attempt at SRI at quantitatively defining an RV response involved reducing the raw transcript to a series of declarative statements called concepts (Targ et al., 1977). Initially, it was decided that a coherent concept should not be reduced to its component parts. For example, a *small red VW car* would be considered a single concept rather than four separate concepts, *small, red, VW,* and *car.* Once a transcript had been "conceptualized," the list of concepts constituted, by definition, the RV response. The analyst rated the concept lists against the sites. Although the response was well defined by this method, no attempt was made to define the target site.

In 1982, a procedure was developed to define both the target and response material (May, 1983). It became evident that before a site can be qualified, the overall remote-viewing goal must be clearly defined. If the goal is simply to demonstrate the existence of the

[2] Although the term *remote viewing* is used throughout this paper, the analysis techniques can easily be applied to any free-response data.

RV phenomenon, then anything that is perceived at the site is important. But if the goal is to gain specific information about the RV process, then possibly specific items at the site are important whereas others remain insignificant.

In 1984, work began on a computerized evaluation procedure (May et al., 1985), which underwent significant expansion and refinement during 1986 (Humphrey et al., 1986). The mathematical formalism underlying this procedure is known as the "figure of merit" (FM) analysis. This method is predicated on descriptor list technology, which represented a significant improvement over earlier "conceptual analysis" techniques, both in terms of "objectifying" the analysis of RV data and in increasing the speed and efficiency with which evaluation can be accomplished. Humphrey's technique, which was based on the pioneering work of Honorton (1975) and its expansion by Jahn, Dunne, and Jahn (1980), was to encode target and response material in accordance with the presence or absence of specific elements.

It became increasingly evident, however, that this particular application of descriptor lists was inadequate in providing discriminators that were "fine" enough to describe a complex target accurately, and unable to exploit fully the more subtle or abstract information content of the RV response. To decrease the granularity of the RV evaluation system, therefore, a new technology would have to allow the analyst a gradation of judgment about target and response features rather than the hard-edged (and rather imprecise) all-or-nothing binary determinations. Requiring an analyst to restrict subjective judgment to single elements rather than to complete responses is consistent with the research reported by Dawes (1988).

A preliminary survey of various disciplines and their evaluation methods (spanning such diverse fields as artificial intelligence, linguistics, and environmental psychology) revealed a branch of mathematics, known as "fuzzy set theory."[3]

*Fuzzy Set Concepts*

Fuzzy set theory was chosen as the focal point of the RV analytical techniques because it provides a mathematical framework for modeling situations that are inherently imprecise. Because it is such an important component in the analysis, a brief tutorial will be presented to highlight its major concepts.

[3] We wish to thank S. James P. Spottiswoode and D. Graff, CE, for directing us to the fuzzy set literature and for many helpful discussions.

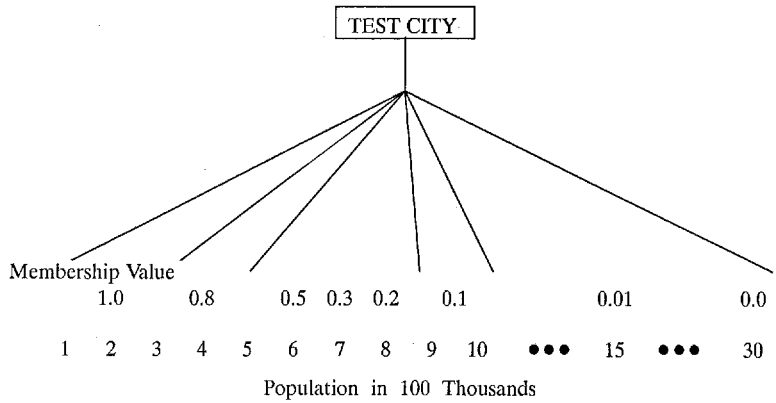


Figure 1. The fuzzy set "kind-of-small" cities.

In traditional set theory (i.e., crisp sets), an element either is or is not a member of a set. For example, the crisp set of cities with population equal to or greater than 1,000,000 includes New York City, but not San Francisco. This set would also *not* include a city with a population of 999,999. The problem is obvious. There is no real difference between cities with populations of 1,000,000 and 999,999, yet one is in the set and the other is not. Humans do not reason this way; therefore, something other than crisp sets is required to capture the subjectivity inherent in RV analysis.

Fuzzy set theory introduces the concept of *degree* of membership. Herein lies the essence of its applicability to the modeling of imprecise concepts. For example, if we consider the size of a city, we might define certain *fuzzy sets*, such as *very small* cities or *kind-of-small* cities. Using *kind-of-small* cities as a fuzzy set example, we might subjectively assert that a city with a population of 100,000 is definitely such a city, but a city with a population of 400,000 is only a little bit like a *kind-of-small* city. As depicted in Figure 1, fuzzy set theory allows us to assign a membership value between 0 and 1 that represents our best subjective estimate as to how much each of the possible city populations embodies the concept *kind-of-small*. In this example, a population of 700,000 assigned a membership value of 0.3.

Clearly, a different set of membership values would be assigned to the populations for the fuzzy sets *very small* cities, *medium* cities, *large* cities, and so forth; a population of 100,000 might receive a value of 0.2 for very small cities, but a value for 1.0 for kind-of-small cities, depending on context, consensus, and the particular

application. These membership values can be obtained through consensus opinion, a mathematical formula, or by several other means. Crisp sets are special cases of fuzzy sets, in which all membership values are either zero or one. By using membership values, we are able to provide manipulatable numerical values for imprecise natural language expressions; in addition, we are no longer forced into making inaccurate binary decisions such as, "Is the city of San Francisco large—yes or no?"

In this example, the crisp set of all cities defines the universal set of elements (USE). The crisp set of cities with populations of one million or more is a subset of USE. The fuzzy sets *very small, kind-of-small, medium,* and *large* cities are fuzzy subsets of USE.

*Universal Set of Elements*

Since targets and the responses will be defined as fuzzy sets, we must specify a USE. The universal set of elements can be quite general and include all aspects of a given target pool, or it can be tailored to a specific experiment to test a given concept (e.g., include only geometric shapes). Since the method of fuzzy set analysis critically depends on the choice of USE, we provide one example that was derived from a target pool used in earlier experiments. What follows is *only* an example of how one might construct a USE. The one we use is not generally applicable to other target pools or other experiments.

We constructed our USE by including a list of features present in photographs from the *National Geographic* magazine with elements obtained from the RV responses in earlier experiments. This USE is presented in Appendix A as the actual coding forms. For the target features, we focused on direct visual elements. (In the general case, other perceptual dimensions can be considered.) In the case of the RV response-derived elements, an effort was made to preserve the vocabulary used by the viewers. Some of the elements, therefore, are either response-dependent or target-dependent or both, whereas others, particularly at the more abstract levels, appear to be more universal across possible USEs.

This universal set of elements is structured in *levels,* ranging from the relatively abstract, information poor (such as vertical lines), to the relatively complex, information rich (such as churches). The current system is structured into seven primary and three secondary levels of elements; the main intent of this structure is to serve as a heuristic device for guiding the analyst into making judicious con-

crete element assignments based on rather abstract commentary. The use of levels is advantageous in that each element level can be weighted separately and used or not, as the case may be. This enables various combinations of levels to be deployed to identify the optimal mix of concrete versus abstract elements. Of course, any such weighting scheme must be determined in advance of any experiment.

The determination as to which elements belonged on which level was made after consideration of two primary factors: (1) the apparent ability of the viewers to be able to resolve certain features, coupled with (2) the amount of pure information thought to be contained in any given element. Some of these "factor one" determinations were based on the combined anecdotal experience of analysts and monitors in the course of either analyzing or conducting numerous RV experiments; some were determined empirically from post hoc analyses of viewers' abilities to perceive various elements in previous experiments.

The "factor two" determinations were made primarily by arranging the elements such that an element at any given level represents the sum of its constituent elements at lower levels. For example, a *port* element (Level 7) could be considered to include *canal* (Level 6) and *partially bounded expanse of water* (Level 5). The world is not a very crisp place and not all its elements are amenable to hierarchical structuring. Certain violations of the "factor two" rule appear, therefore, throughout the USE example. It should be noted, however, that some of the more glaring violations were largely driven by the "factor one" determinations (i.e., the viewers' abilities to discern certain elements) enumerated above.

To emphasize once again, it is very important to realize that this universal set of elements was constructed to match our particular special targets, viewers, and requirements. They are shown here to illustrate the procedure. Any particular application of fuzzy set technology to the analysis of free-response material requires an a priori construction of an individualized, and improved, USE specific to the target pool and the goals of the experiment.

*Target Fuzzy Sets*

Each target is defined as a fuzzy set constructed by assigning a membership value to each of the elements in the USE (see Appendix A). In general, membership values can vary continuously on the interval [0,1]. In this application they represent human judgment

and, thus, were constrained to vary in steps of 0.1. In addition, they must represent the perceptual dimension used to construct the USE. In our example, membership values were assigned to each element for each of the targets, according to a consensus (on an element-by-element basis) reached by three analysts. This approach was used to mitigate the potential influence of any single coder's biases and idiosyncrasies. A numerical assignment, $\mu$ ($0 \leq \mu \leq 1$, in steps of 0.1), was made for each element in response to the following question: How visually important is this element to this photograph?

Encoded by this method, the fuzzy sets served as a formal definition of the targets for the analysis. It should be noted that our USE defined targets in terms of visual importance.[4] If other dimensions are of interest (e.g., conceptual, functional, allegorical), the USE would have to be revised to incorporate them.

In an actual experimental series, it is critical that the target fuzzy sets be defined by analysts *before* the series begins. Because of the potential information leakage owing to bias on the part of the analyst, it is an obvious mistake to attempt to define the target fuzzy set on a target-by-target basis in real time or post hoc.

*Response Fuzzy Sets*

To define RV response fuzzy sets, membership values $\mu$ are assigned for each element in the USE by asking: To what degree am I (the analyst) convinced that this element is represented in this response? For example, if a response explicitly states "water," then the membership value for the water-element should be 1. If, however, the response is a rough sketch of what might be waves, then the membership value for the water-element might be only 0.3, depending on the specificity of the drawing. This definition of membership value is quite general and can be used in most applications.

In our example, responses were coded according to this definition (but still using the USE in Appendix A). The assigned $\mu$'s for the targets and responses were one-digit fuzzy numbers on the interval [0,1] (e.g., 0.1, 0.2, 0.3, etc.). In some rare cases, two-digit assignments (e.g., 0.05, 0.15, 0.25, 0.35, etc.) were made; any finer assignments, however, were deemed to be meaningless. Thus, the response was defined as its fuzzy subset of the USE.

[4] Implied visual importance was ignored. For example, in a photograph of the Grand Canyon that did not show the Colorado River, water, river, and so on would be scored as zero. By definition the target was only what was visible in the photograph.

In an actual experimental series, each response fuzzy set is created by analysts who are blind to the intended target.

*Fuzzy Set Definition of Figure of Merit*

Once the fuzzy sets that define the target and the response have been specified, the comparison between them to provide a figure of merit (FM) is straightforward. In previous work (Humphrey et al., 1986), we have defined *accuracy* as the percent of the target material that was described correctly by a response. Likewise, we have defined *reliability* (of the viewer) as the percent of the response that was correct. The FM is the product of the two; to obtain a high FM, a response must be a comprehensive description of the target and be devoid of inaccuracies. The mathematical definitions for accuracy and reliability for the *j*th target/response pair are as follows. Let $\mu_k(R_i)$ and $\mu_k(T_j)$ be the membership values for the *k*th element in USE for the *i*th response and the *j*th target, respectively. Then the accuracy and reliability for the *i*th response applied to the *j*th target are given by:

$$\text{accuracy}_{ij} = a_{ij} = \frac{\sum_k W_k \min\{\mu_k(R_i), \mu_k(T_j)\}}{\sum_k W_k \mu_k(T_j)},$$

$$\text{reliability}_{ij} = r_{ij} = \frac{\sum_k W_k \min\{\mu_k(R_i), \mu_k(T_j)\}}{\sum_k W_k \mu_k(R_i)}$$

where the sum over $k$ is called the *sigma count* in fuzzy set terminology, and is defined as the sum of the membership values. We have allowed for the possibility of weighting the membership values with weights $W_k$ in order to examine various level/element contributions to the FM. The index, $k$, ranges over the entire USE.

For the above calculation to be meaningful, the $\mu$'s for the targets must be similar in meaning to the $\mu$'s for the responses. As we noted above, in our definition of the membership values, this is not the case. The target $\mu$'s represent the visual importance of the element relative to the scene, and the response $\mu$'s represent the degree to which an analyst is convinced that the element is represented in the response regardless of its relevance to that response.

With advanced viewers it might be possible to change the definition of the response $\mu$'s to match the definition of the target $\mu$'s. In that case, the viewer must not only recognize that an element is

present in the target, but must also provide information as to how visually important it is. This ability is currently beyond the skill of most novice viewers. Alternatively, we have opted to modify the target $\mu$ definition by using the fuzzy set technique of $\alpha$-cuts. In our example, an $\alpha$-cut is a way to set a threshold for visual importance. All target elements possessing that threshold value or higher are considered to be full members of the target set. In fuzzy set parlance, an $\alpha$-cut converts a fuzzy set to a crisp one. The result is that the target set is now devoid of detailed visual information: a potential target element is either present or absent in the target set, regardless of its actual visual importance. Even with this conceptual change in the target definition, the FM formalism described above remains applicable, because a crisp set can be considered as a fuzzy set with all membership values equal to 0 or 1. It is important to recognize that the $\alpha$-cut is only applied to the target set; the response set remains fuzzy.

*Assessment of Quality of the Remote Viewing*

It is difficult to arrive at a general assessment of how well a given response matches a specified target. The ideal situation is to obtain some absolute measure of goodness of match. Although the FM is an approximation to this measure, it is impossible to assess the likelihood of a particular FM value because it requires knowledge of the viewer's *specific* response bias for the session. It is possible to determine general response biases (May et al., 1985), but that knowledge is only useful on the average. For example, a viewer may love rock climbing and may spend most of his free time involved in that activity. Thus, the general response bias would probably entail aspects of mountains, rocks, ropes, and so forth. Suppose, however, that the viewer spent the evening previous to a given RV session on a romantic moonlight sail on San Francisco Bay. For this specific RV session, the response bias might include romantic images of the moonlit water, lights of the city, and bridges.

The current solution to the problem is to provide a *relative* assessment of FM likelihood. A relative assessment addresses the following question: "How good is the response matched against its intended target, when compared to all possible targets that could have been chosen for the session?" This is not ideal, since the answer depends on the nature of the remaining targets in the pool. An example of the worst-case scenario illustrates the problem. Suppose

that the target pool consisted of 100 photographs of waterfalls, and the viewer gave a near-perfect description of a waterfall. (We assume that this description is not fortuitous.) An absolute assessment of the resulting FM should be good, whereas a relative assessment will be low. The worst-case scenario can be avoided, to a large degree, by carefully selecting the target pool. (See the later section "A Quantitative Definition of Target Orthogonality.")

To provide a relative assessment of the likelihood of a given FM, we define the score for one session to be the number of targets, $n$, out of a total, $N$, that have an FM equal to or higher than the FM achieved by the correct match.[5] The answer to the question: "Given this response, what is the probability of selecting a target that would match it as well as or better than the target selected?" is $n/N$.

Consecutive RV responses by the same viewer are not statistically independent, nor can the responses be considered to be random in any sense. The statistically independent random element in the session is the target. Since targets are selected with replacement, under the null hypothesis of no psi, the collection of scores derived over a series of $m$ trials constitutes a set of independent random variables, each with a discrete uniform distribution. Under the null hypothesis, the mean chance expectation for the score in each session is given by $(N + 1)/2$ and the variance is given by $(N^2 - 1)/12$. If $K$ is the sum of scores from a series of remote viewings, then the probability of $K$, under the null hypothesis, can be obtained from the exact distribution for the sum of ranks given by Solfvin, Kelly, and Burdick (1978):

$$p(K \text{ or less}) = \frac{1}{N^m} \sum_{a=m}^{K} \sum_{b=0}^{N} (-1)^b \binom{m}{b} \binom{a - bN - 1}{m - 1}. \qquad (1)$$

If $m$ is large, then the sum-of-ranks distribution is approximately normal and $K/m$ has a mean of $(N + 1)/2$ and a variance of $(N^2 - 1)/12m$. Thus, a $z$ score can be computed from:

$$z(K \text{ or less}) = \frac{0.5(N + 1) - \frac{K}{m}}{\sqrt{\frac{N^2 - 1}{12m}}}. \qquad (2)$$

[5] $N$ must be the size of the target pool from which each target was randomly selected, and for this theoretical discussion, we assume no ties.

*Ground Truth*

To determine whether the new analytical approach was effective, a standard had to be developed against which it could be measured. It was determined that this standard—known as "ground truth"—should consist of a "real-world" normalized consensus about the degree of correspondence between RV responses and their intended targets.

To achieve this objective, we presented analysts (chosen from the general SRI staff) with the same test case of six remote-viewing responses and their associated targets. The test case was the data from a single viewer (177) taken from an experimental series in a 1986 photomultiplier tube experiment (Hubbard, May, & Frivold, 1987). The responses (i.e., two to five pages of rudimentary drawings with some associated descriptive words) were fairly typical of novice viewer output and represented a broad range of response quality. The targets consisted of six photographs of outdoor scenes selected from a *National Geographic* magazine target pool of 200. Thus, this data set was ideally suited for an analysis testbed. Appendix B contains the "best" and "worst" trials (Sessions 9005 and 9004, respectively) from this series in the form of their responses, their intended targets, and their fuzzy set encodings (see the next section).

Each analyst was asked individually for his subjective judgment about the degree of correspondence between the remote-viewing responses and their respective intended targets. The "degree of correspondence" was purposely undefined; the analysts had to formulate their own criteria. The only information provided was that responses typically begin with small bits of information and eventually culminate in a composite drawing at the end. Appendix C contains the coding form that was used to obtain "ground truth."

Each analyst was instructed to examine all of the responses and their intended targets. Then, on a session-by-session basis, he was asked: (1) to assess the degree of correspondence between the remote-viewing response and its intended target, and (2) to register this correspondence assessment by making a vertical hash mark across a 10-cm scale ranging from "none" to "complete."

To perform the ground truth analysis, distance measurements were taken from the left end point of each scale to the vertical slash mark for each assessment. Let the distance obtained for the $k$th ses-

sion from the $j$th analyst be given by $x_{j,k}$. To account for analysts' biases, the $x_{j,k}$ were normalized by a $z$ transformation,

$$z_{j,k} = \frac{x_{j,k} - \mu_j}{\sigma_j},$$

where $\mu_j$ and $\sigma_j$ are the mean and standard deviation of the $j$th analyst's distance scores, $x_{j,k}$. The effect of this transformation is to convert an analyst's absolute subjective opinion to a relative one. For the $j$th analyst, the largest $z_{j,k}$ indicates that the degree of correspondence for response/target $k$ is higher than *any other pair* in the series. It does *not* indicate overall quality. This type of transformation was necessary since we wished to combine the assessments from a number of different analysts.

To combine the assessments across analysts, we computed the mean $z$ score for each response/target pair, $k$, as:

$$z_k = \frac{1}{N_a} \sum_{j=1}^{N_a} z_{j,k},$$

where $N_a$ is the number of analysts. The number of analysts was determined by the data. For the best response/target pair (i.e., session 9005, $k = 5$) we computed the percent change of $z_5$ for every additional analyst. When the addition of two new analysts produced consecutive changes of less than 2%, the process was considered complete. For this data set, 37 analysts were required before this condition was met. Figure 2 shows the normalized mean for each target/response pair, and represents a relative assessment of remote-viewing quality. These means constitute the basis for the ground truth against which the fuzzy set technique was measured. We recognize that this definition of ground truth is based on global decisions and may not be most optimal (Dawes, 1988).

*Results of the Fuzzy Set Analysis*

To effect a meaningful comparison between ground truth and the figure of merit analysis, we also analyzed the same RV series that served as the ground truth set by the fuzzy set figure of merit method. The fuzzy set membership values ($\mu$'s) for the six targets and six responses were consensus coded by five analysts ranging from expert to novice. A typical spread of $\mu$ assignments was $\pm$ 0.1 with an occasional outlier. Some of the elements were vigorously debated until a consensus was reached. Accuracies, reliabilities, and

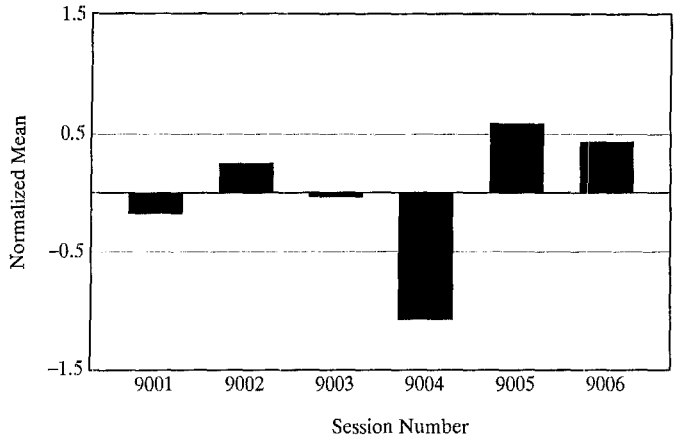


Figure 2. Normalized mean for each target/response pair.

figures of merit were calculated for each target/response pair (Table 1). It should be noted that the encoding was a post hoc exercise, but because the assignment for each element in the USE had to be defended before a consensus was reached, the FMs shown in Table 1 constitute reasonable estimates of their "blind" equivalents. Appendix B shows the target and response elements that were scored from the universal set (see Appendix A) for Sessions 9004 and 9005. As an example of the fuzzy calculation, Appendix B also shows the re-

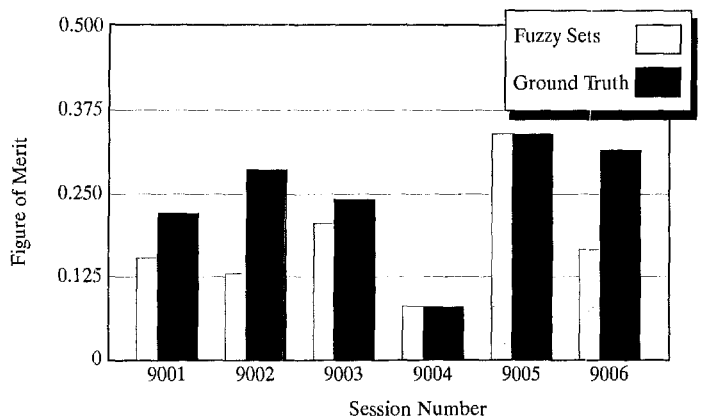


Figure 3. Comparison with ground truth.

TABLE 1
FUZZY SET QUANTITIES FOR "GROUND TRUTH" SERIES

| Session | Accuracy | Reliability | Figure of merit | Rank | Fractional rank |
|---|---|---|---|---|---|
| 9001 | .317 | .484 | .153 | 80 | .403 |
| 9002 | .273 | .477 | .130 | 103 | .515 |
| 9003 | .358 | .571 | .205 | 31 | .155 |
| 9004 | .212 | .379 | .080 | 142 | .713 |
| 9005 | .573 | .594 | .340 | 3 | .015 |
| 9006 | .298 | .555 | .165 | 13 | .068 |

sults of the target α-cut, the fuzzy intersection, and the accuracy, reliability, and figure of merit for Session 9005. Table 1 also shows the absolute and relative ranks from a target pool of 200. To determine the absolute rank for each session, we calculated figures of merit for all 200 targets in the pool and placed them in numerical order from the largest to the smallest. The absolute rank is just the position (from the top) of the FM corresponding to the intended target. Ties were resolved by choosing the next larger integer rank number to the centroid of the ties. The fractional rank number can be considered a $p$ value for an individual session and is equal to the absolute rank/200. Using Equation 1, the overall $p$ value for the combined six trials is .052 ($N = 200$, $K = 372$, $m = 6$). Using the approximation (Equation 2), we compute $z = 1.633$, $p \leq .05$, to demonstrate that for six trials, the approximation in reasonable. For completeness, we compute the effect size ($r = 0.67$).

To compare the results of the fuzzy set analysis with those of the ground truth, we linearly renormalized the ground truth figures to be within the interval [0,1] and to possess the same maximum and minimum. As can be seen from Figure 3, the results from the fuzzy set analysis system parallel those obtained by a consensus of the 37 analysts each making a subjective assessment of the matches.

These results imply that the combination of (1) the structure of the USE (i.e., the linguistic hierarchical structure), (2) the fuzzy set mathematics, and (3) a consensus approach to assessing the fuzzy sets themselves provided a reasonable representation of the subjective scoring of the same data by a large number of individuals.

*A Quantitative Definition of Target Orthogonality*

It is often of interest to define how similar or dissimilar targets are to each other. For example, free-response experiments like the

ganzfeld often use target packets, with the unselected targets in a packet serving as decoys for judging. Assigning potential targets to packets would be easier with some measure of target orthogonality.

Target definition for the purposes of this mode of analysis is exactly the same as the one described (i.e., a given target is defined by its fuzzy subset of the USE, which has been coded to reflect the visual importance of each target element). The average number of elements, of the total of 131, that was assigned a nonzero value for the targets in our pool of 200 was approximately 37, indicating that the fuzzy set representation of the target pool is rich in visual information. We used this information to determine the degree to which the target set contains visually similar targets.

It is beyond the scope of this paper to describe the extensive work in the literature seeking to find algorithmic techniques that mimic human assessments of visual similarity. One recent article describes techniques similar to the one we used (Zick, Carlstein, & Budescu, 1987).

We begin by defining the simlarity between target $i$ and target $j$ to be a normalized fuzzy set intersection between the two target sets:

$$S_{ij} = \frac{\left(\sum_k W_k \min\{\mu_k(T_i), \mu_k(T_j)\}\right)^2}{\sum_k W_k \mu_k(T_i) \sum_k W_k \mu_k(T_j)},$$

where the index $k$ ranges over the entire USE. We have allowed for the possibility of weighting the membership values with weights $W_k$ to examine various level/element contributions to the target similarities.

For $N$ targets, there are $N(N-1)/2$ unique values (19,900 for $N = 200$) of $S_{ij}$. The values $i$ and $j$ that correspond to the largest value of $S_{ij}$ represent the two targets that "look" most similar. Suppose another target $m$ is chosen and $S_{m,i}$ and $S_{mj}$ are computed. If both of these values are larger than $S_{m,n}$ (for all $n$ not equal to $i$ or $j$), then target $m$ is assessed to be most similar to the pair $ij$. The process of grouping targets based on these similarities is called *cluster analysis.*

Using this process, 200 targets were grouped into 19 clusters, such that the targets are similar within a cluster, and dissimilar between clusters. Table 2 provides an overview of the 19 clusters found from the total analysis of the 200 targets. Some of the names appear to be quite similar, but, in fact, these sets are visually quite distinctive. Figure 4 shows the graphic output of a single cluster in

TABLE 2
NAMES OF THE 19 CLUSTERS

| No. | Name | No. | Name |
|---|---|---|---|
| 1 | Flat towns | 11 | Cities with prominent geometries |
| 2 | Waterfalls | 12 | Snowy mountains |
| 3 | Mountain towns | 13 | Valleys with rivers |
| 4 | Cities with prominent structure | 14 | Meandering rivers |
| 5 | Cities on water | 15 | Alpine scenes |
| 6 | Desert/water interfaces | 16 | Outposts in snowy mountains |
| 7 | Deserts | 17 | Islands |
| 8 | Dry ruins | 18 | Verdant ruins |
| 9 | Towns on water | 19 | Agricultural scenes |
| 10 | Outposts on water | | |

detail. A much more complex—and visually difficult to understand—graph is generated for the full cluster analysis and is not included here; this smaller subset, therefore, has been chosen to be illustrative of the whole analysis. All targets in this particular sample cluster are islands; the island in each photograph is visible in its entirety. Except for one outlier (i.e., a hexagonal building covering an island), the islands fall into two main groups (i.e., with and without

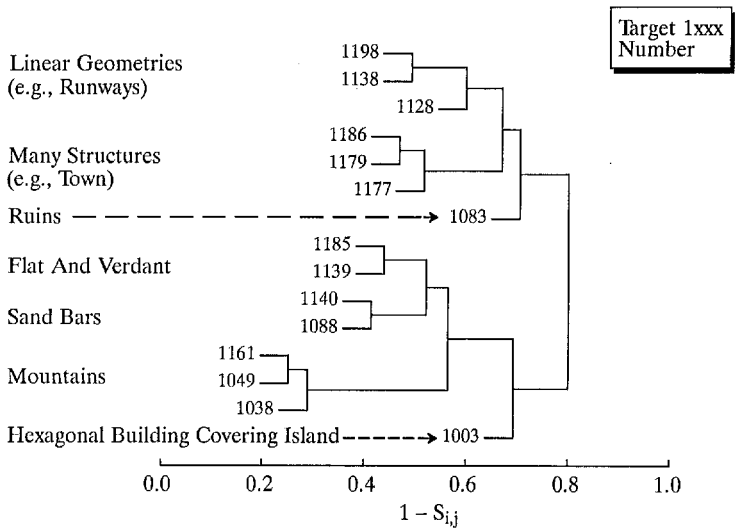


Figure 4. Cluster analysis of island targets.

manmade elements). The natural islands include three similar mountain islands, two sandbars, and two flat verdant islands.

Using cluster analysis in conjunction with fuzzy set analysis provides for a quantitative definition of sets of targets that are similar to each other within a cluster, but visually different across clusters. Orthogonal clusters can be used to provide visual decoy targets for traditional rank-order judging.

*Recommendations and Conclusions*

To apply the analysis in its present form to a long RV series is quite labor intensive and, from the results shown in Figure 3, is most likely not justified since this fuzzy set technique approximates human assessment. As we stated in the introduction, however, we are providing only a progress report of ongoing research. Because of the decision concepts described in Dawes (1988) and the obvious benefits of an automated evaluation system, the effort to improve what was described in this paper is certainly justified. The procedure can be used "as is" to improve and quantify target orthogonality.

Several future research areas are suggested to improve the techniques described in this paper. The use of both inter- and intra-level weighting factors needs to be examined systematically. In the analysis described above, all levels and elements were accorded equal weight. The ideal goal would be to determine the optimal weighted mix of abstract versus concrete elements, as a means to achieving the following objectives:

1. Refinement of the cluster analysis for targets, in an effort to simulate, as closely as possible, what is meant by "visual similarities" between targets.

2. Refinement of the analysis of responses, in an effort to achieve even greater correlations between the fuzzy set figure of merit analysis and various forms of ground truth.

Another area that requires examination in some detail is the USE and the hierarchical nature of its structure. It is probable that some elements are more appropriate than others; furthermore, they might be more effectively structured in a semantic network as opposed to a true hierarchy. If a hierarchical structure is retained, then some attention must be paid to the formulation of logical consistency rules that govern element use. This would include numeri-

cal relationships governing the membership values ($\mu$'s) of higher-order elements (e.g., *port*) vis-à-vis the combined value of their constituent parts (e.g., *city, river, boats, jetties, commercial*).

One inadequacy of the system is that it atomizes conceptual "units." For example, if the response element is *red box*, information is lost in separating *red* from *box*. Current research in fuzzy set theory indicates that fuzzy aggregates of fuzzy elements—"fuzzy sets of fuzzy sets"—are mathematically complex but possible. Some effort should be made to determine whether this technology could be implemented as a means to capturing the information content of the RV response with greater accuracy.

For the visual analysis, research into visual similarities between pictures of natural scenes may serve as a potential refinement tool. The aim here would be to enhance the visual orthogonality of rank-order analysis decoy targets as much as possible. Experiments in normal perception of similarities would assist in determining whether scenes are perceived as similar because of their low-level geometries, concrete elements, or some combination of factors. The ultimate aim would be to refine the target cluster analysis such that it closely simulates ground truth representations of orthogonality.

## Appendix A
## Coding Forms for the Universal Set of Elements

The following coding forms illustrate the use of a universal set of elements (USE) that matched our particular special targets, viewers, and requirements. We constructed our USE by including a list of features present in photographs from the *National Geographic* with elements obtained from the remote-viewing responses in earlier experiments.

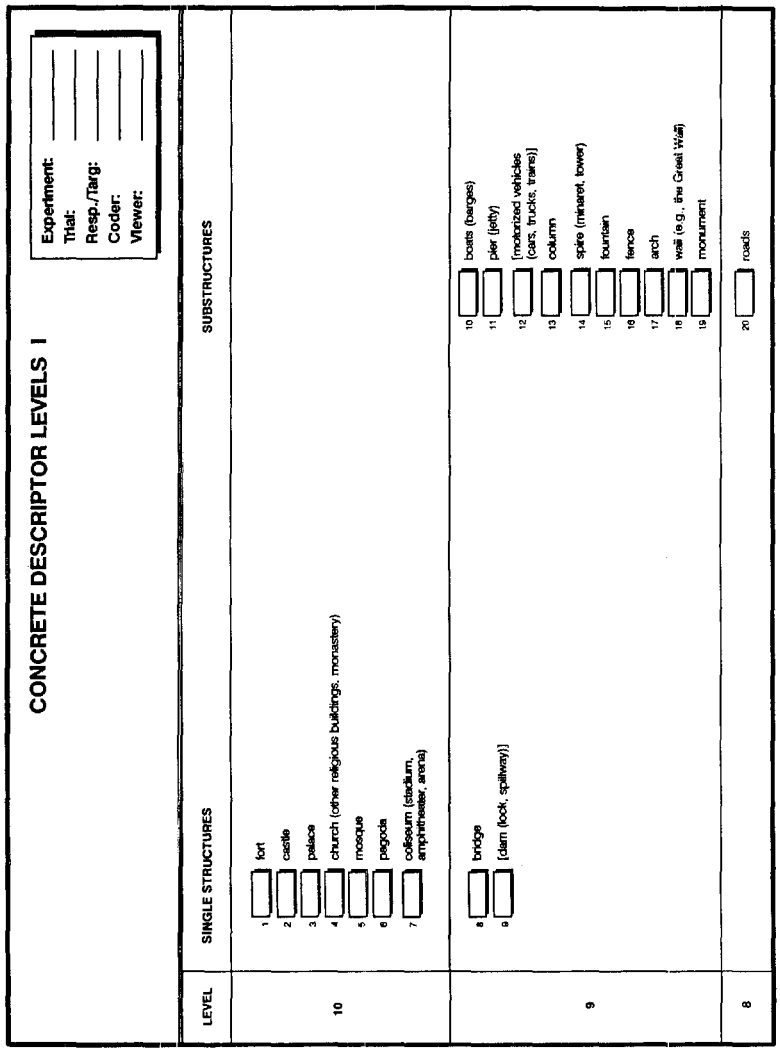


Figure A1. Coding form.

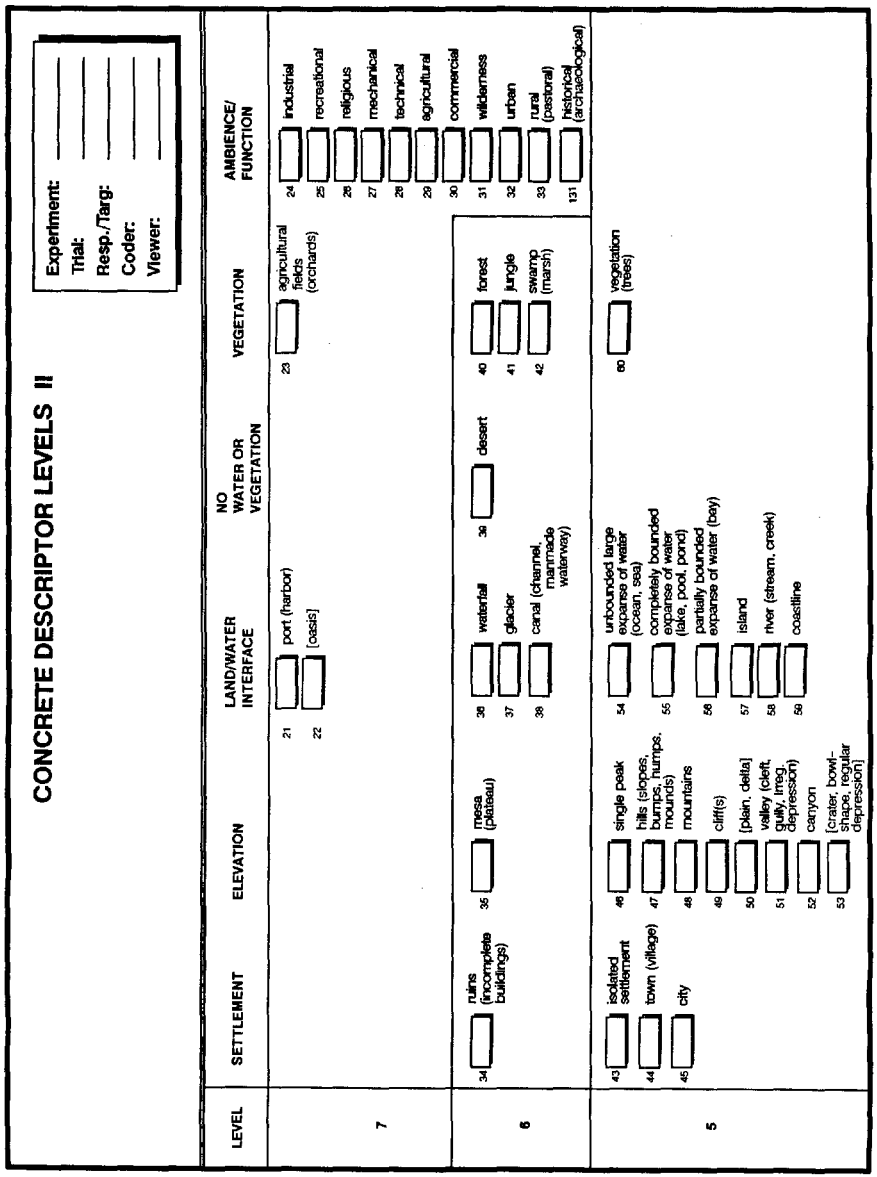


Figure A2. Coding form.

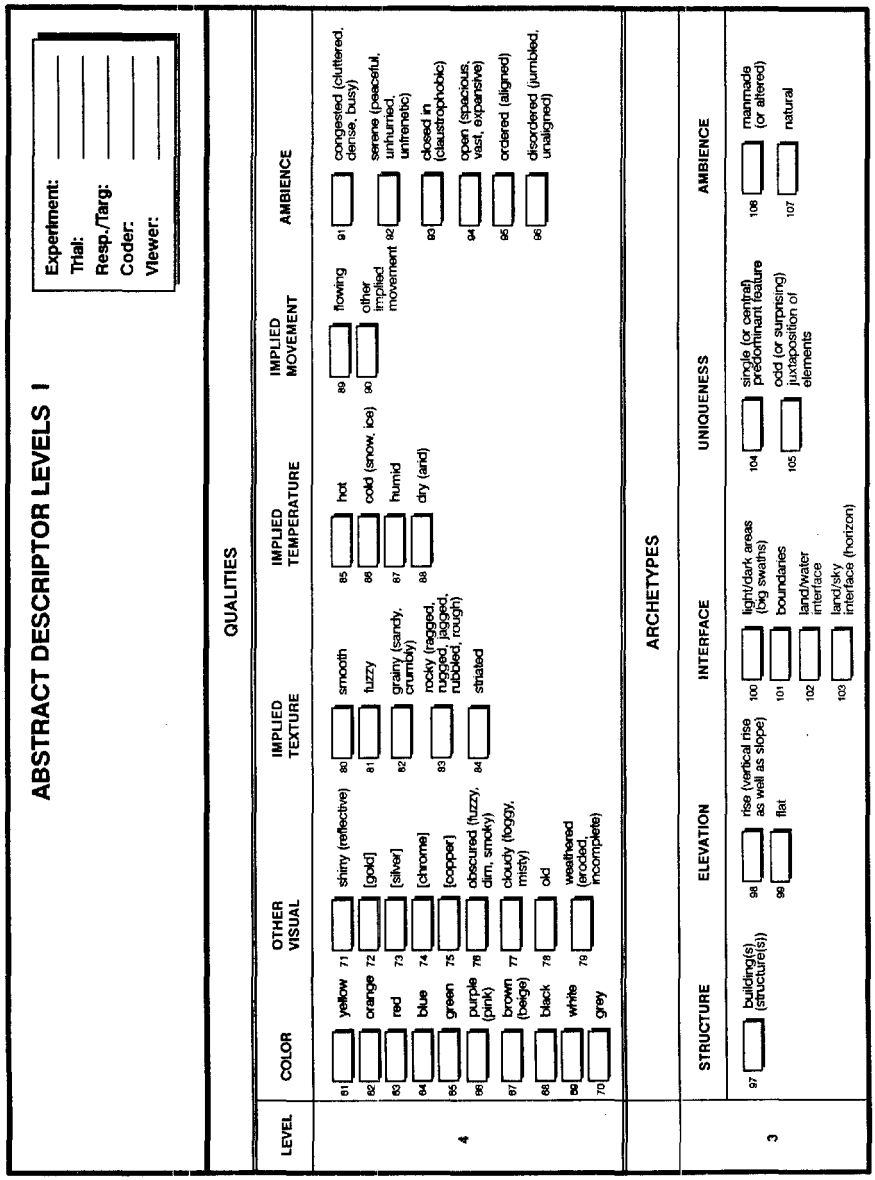


Figure A3. Coding form.

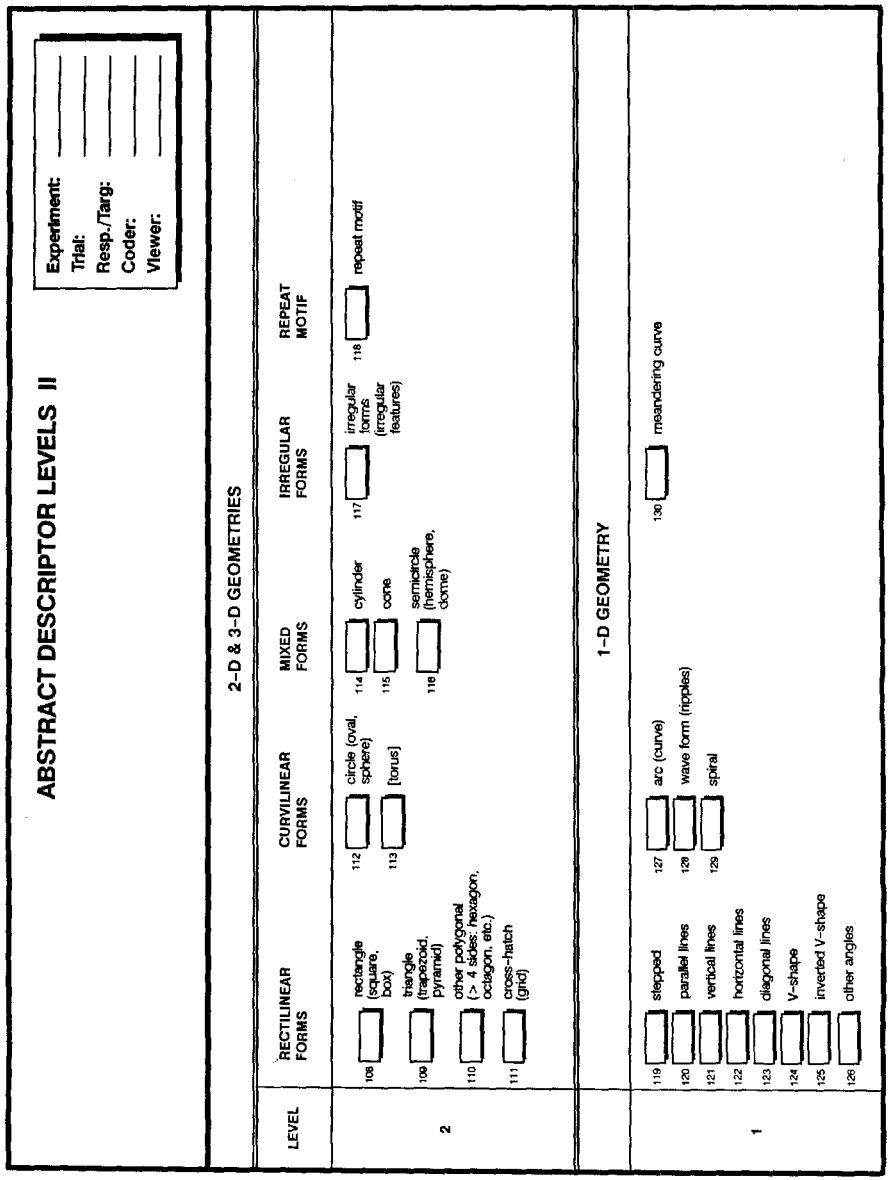


Figure A4. Coding form.

## Appendix B
## Fuzzy Set Analysis Testbed

The following pages show the targets, responses, and analysis for two remote-viewing trials.

Figure B1. Target for Session 9004.

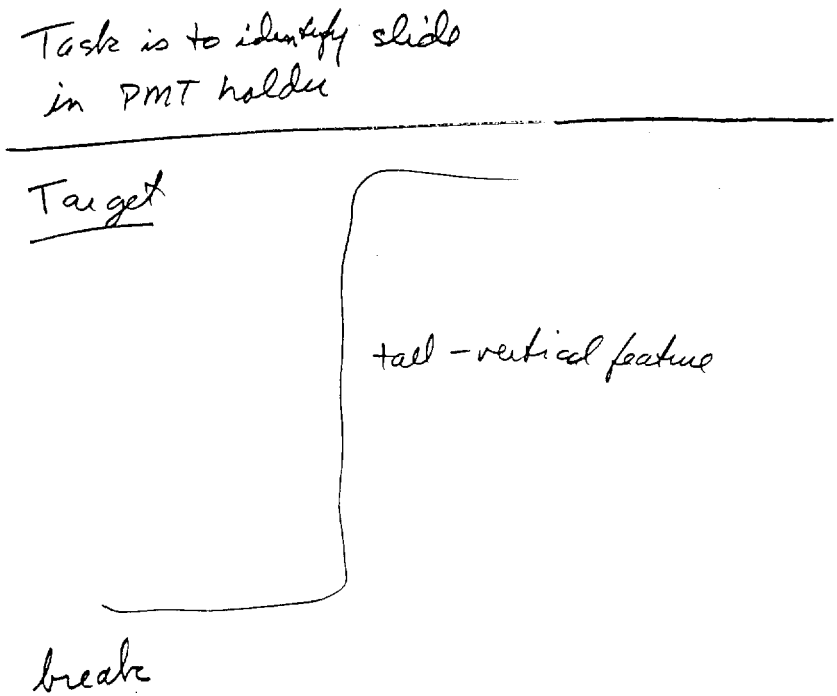


Figure B2. Page one of the response (Session 9004, Target 1094)

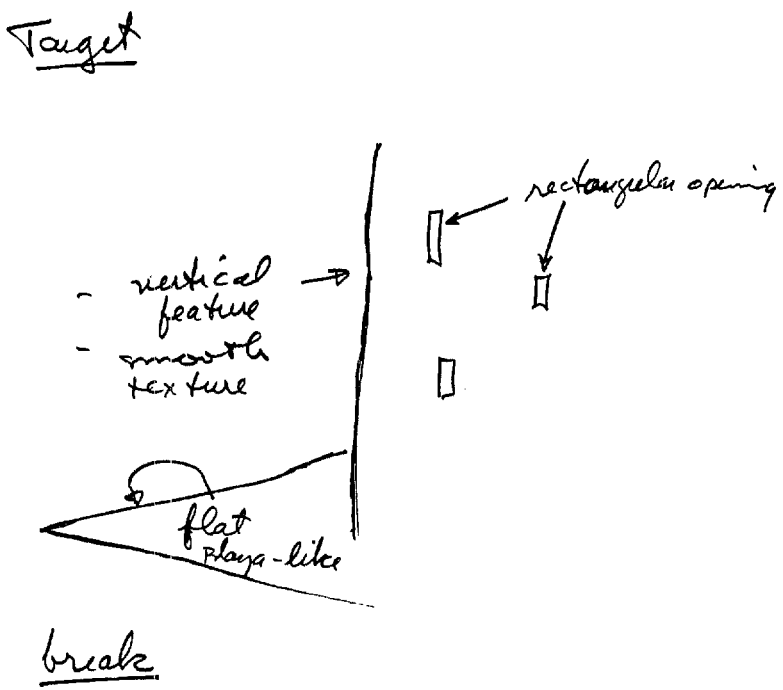


Figure B3. Page two of the response (Session 9004, Target 1094).

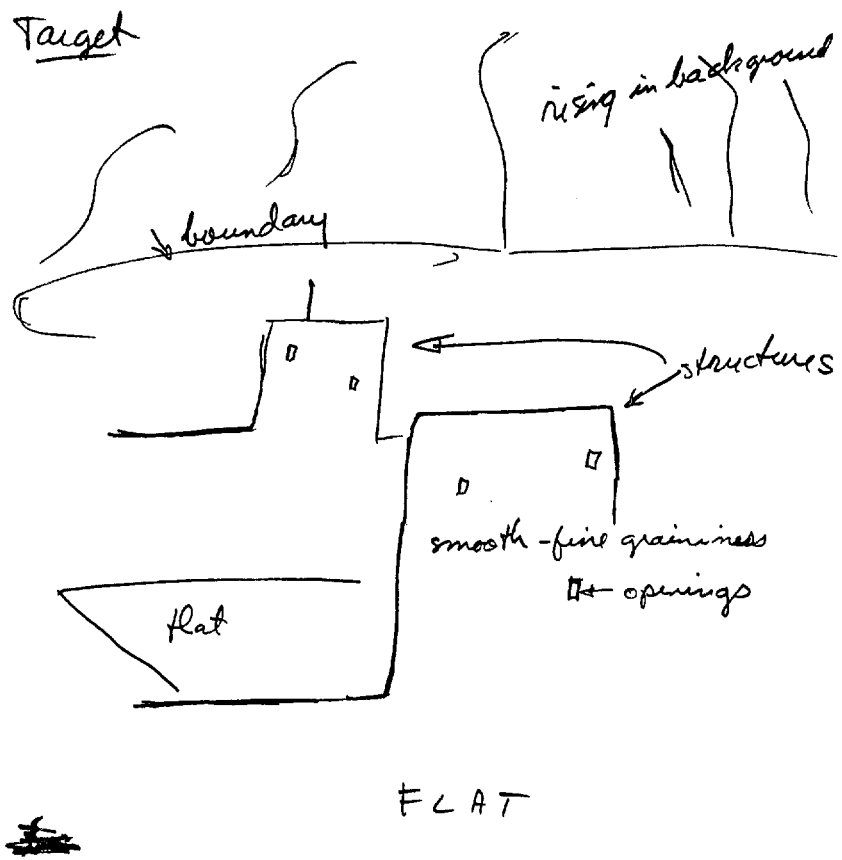


Figure B4. Page three of the response (Session 9004, Target 1094).

TABLE B1
TARGET–RESPONSE 9004

| Element | Name | Target | Response |
|---|---|---|---|
| 20 | Roads | 0.30 | 0.00 |
| 23 | Agricultural fields | 0.05 | 0.00 |
| 32 | Urban | 0.00 | 0.50 |
| 33 | Rural, pastoral | 0.60 | 0.50 |
| 44 | Town, village | 0.00 | 0.50 |
| 45 | City | 0.00 | 0.40 |
| 46 | Single peak | 0.70 | 0.00 |
| 47 | Hills, slopes, bumps, mounds | 0.10 | 0.40 |
| 48 | Mountains | 0.00 | 0.60 |
| 49 | Cliffs | 0.00 | 0.10 |
| 60 | Vegetation, trees | 0.30 | 0.00 |
| 64 | Blue | 0.50 | 0.00 |
| 65 | Green | 0.30 | 0.00 |
| 69 | White | 0.10 | 0.00 |
| 70 | Grey | 0.20 | 0.00 |
| 76 | Obscured, fuzzy, dim, smoky | 0.20 | 0.00 |
| 77 | Cloudy, foggy, misty | 0.20 | 0.00 |
| 79 | Weathered, eroded, incomplete | 0.00 | 0.10 |
| 80 | Smooth | 0.00 | 1.00 |
| 81 | Fuzzy | 0.20 | 0.00 |
| 82 | Grainy, sandy, crumbly | 0.20 | 1.00 |
| 90 | Other implied movement | 0.20 | 0.00 |
| 91 | Congested, cluttered, busy | 0.10 | 0.30 |
| 92 | Serene, peaceful, unhurried | 0.40 | 0.00 |
| 93 | Closed in, claustrophobic | 0.00 | 0.10 |
| 94 | Open, spacious, vast | 0.60 | 0.00 |
| 95 | Ordered, aligned | 0.00 | 0.40 |
| 97 | Buildings, structures | 0.00 | 1.00 |
| 98 | Rise, vertical rise, slope | 0.60 | 1.00 |
| 99 | Flat | 0.30 | 1.00 |
| 100 | Light/dark areas | 0.10 | 0.00 |
| 101 | Boundaries | 0.30 | 1.00 |
| 103 | Land/sky interface | 0.50 | 0.00 |
| 104 | Single predominant feature | 0.60 | 0.00 |
| 105 | Odd juxtaposition, surprising | 0.30 | 0.00 |
| 106 | Manmade, altered | 0.20 | 0.80 |
| 107 | Natural | 0.70 | 0.20 |
| 108 | Rectangle, square, box | 0.00 | 1.00 |
| 109 | Triangle, pyramid, trapezoid | 0.60 | 0.00 |
| 115 | Cone | 0.60 | 0.00 |
| 117 | Irregular forms | 0.00 | 0.20 |
| 118 | Repeat motif | 0.10 | 0.60 |
| 119 | Stepped | 0.10 | 0.70 |
| 120 | Parallel lines | 0.10 | 0.00 |
| 121 | Vertical lines | 0.10 | 1.00 |
| 122 | Horizontal lines | 0.10 | 0.00 |
| 123 | Diagonal lines | 0.40 | 0.00 |
| 125 | Inverted V-shape | 0.70 | 0.00 |
| 126 | Other angles | 0.00 | 0.10 |

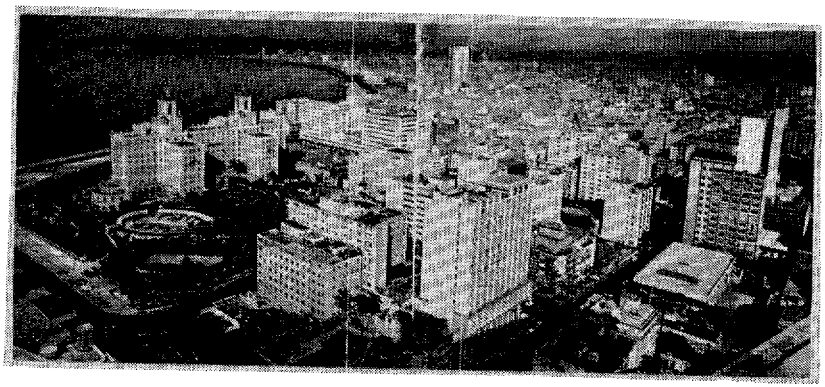

Figure B5. Target for Session 9005.

Task is to describe slide
target in PS 345

Target

interface
cursing
water
Land
break

Figure B6. Page one of response (Session 9005, Target 1005).

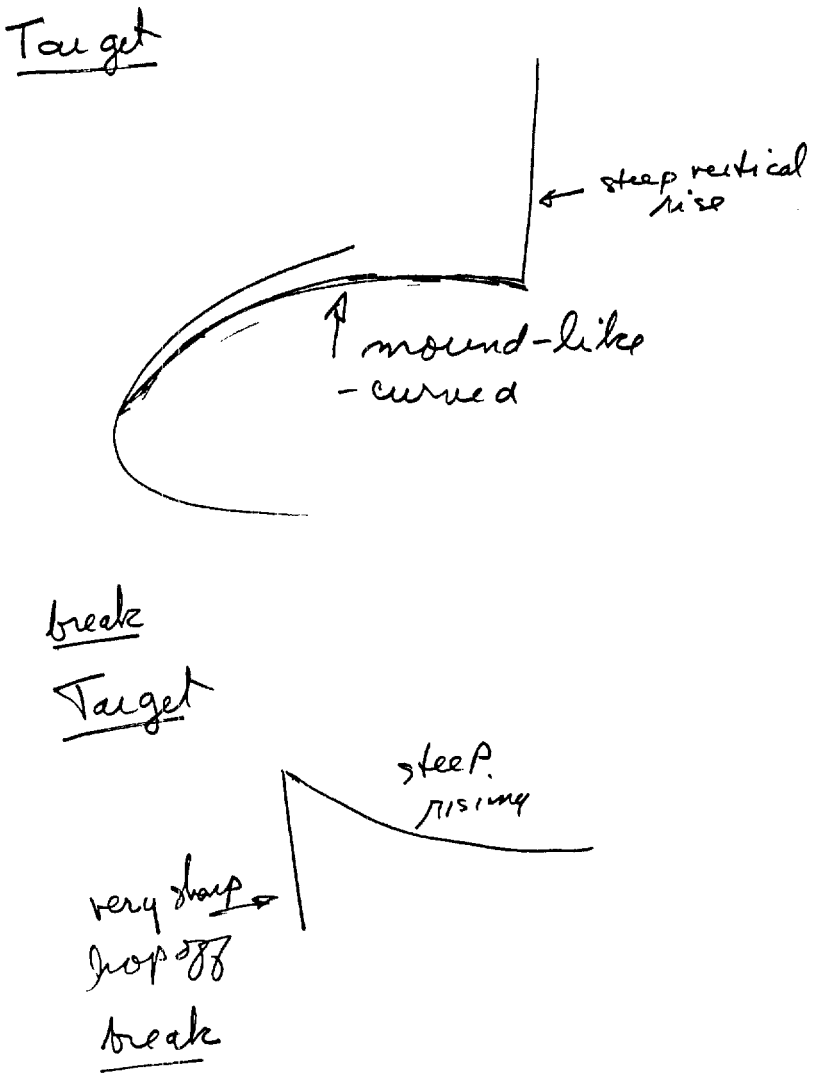


Figure B7. Page two of response (Session 9005, Target 1005).

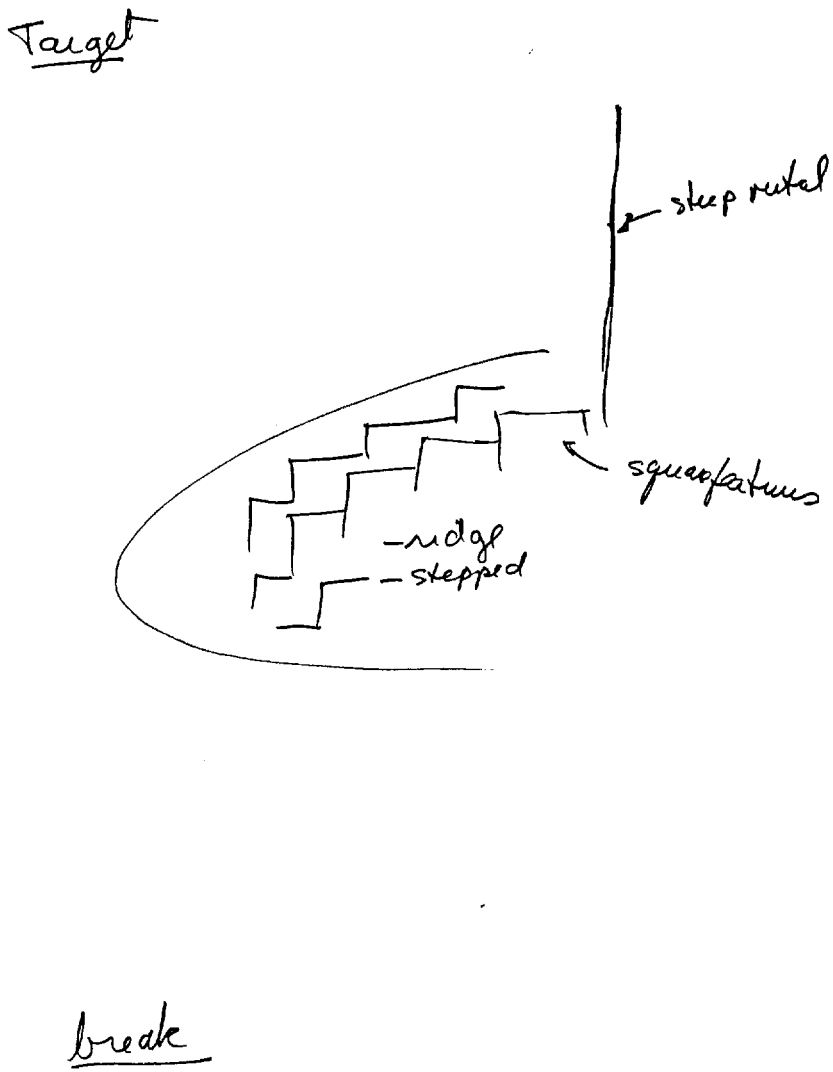


Figure B8. Page three of response (Session 9005, Target 1005).

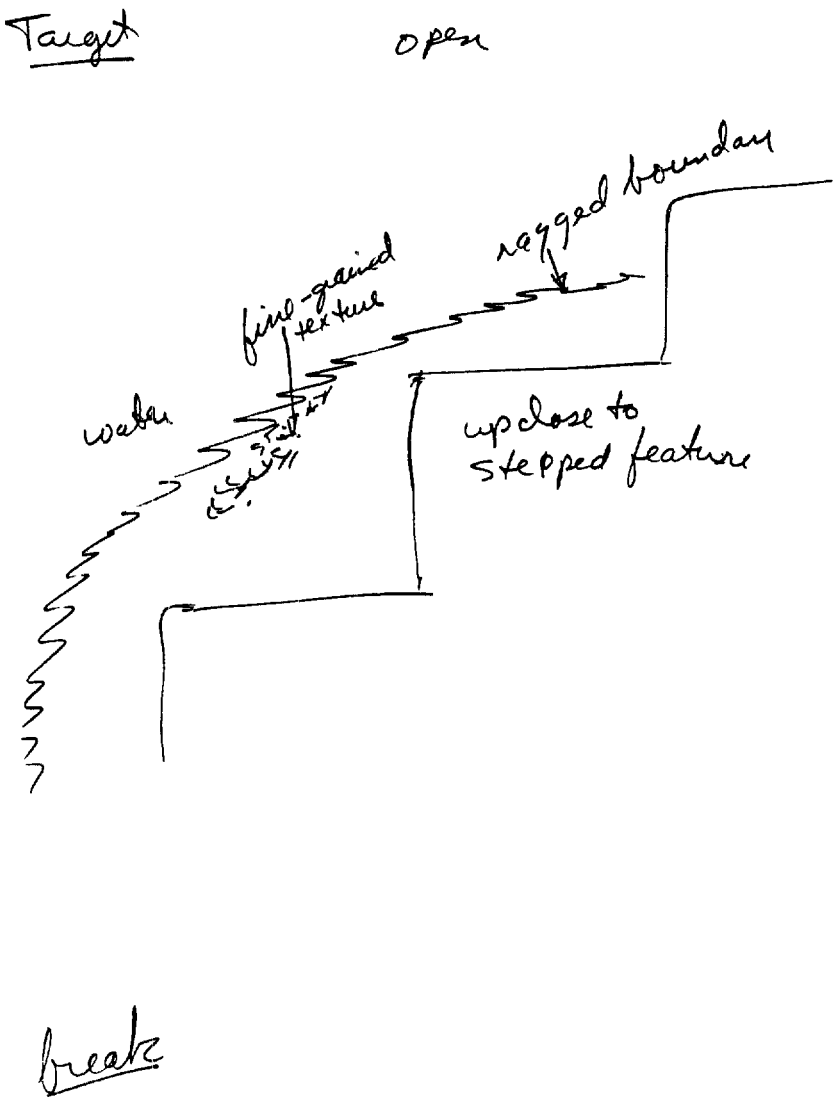


Figure B9. Page four of response (Session 9005, Target 1005).

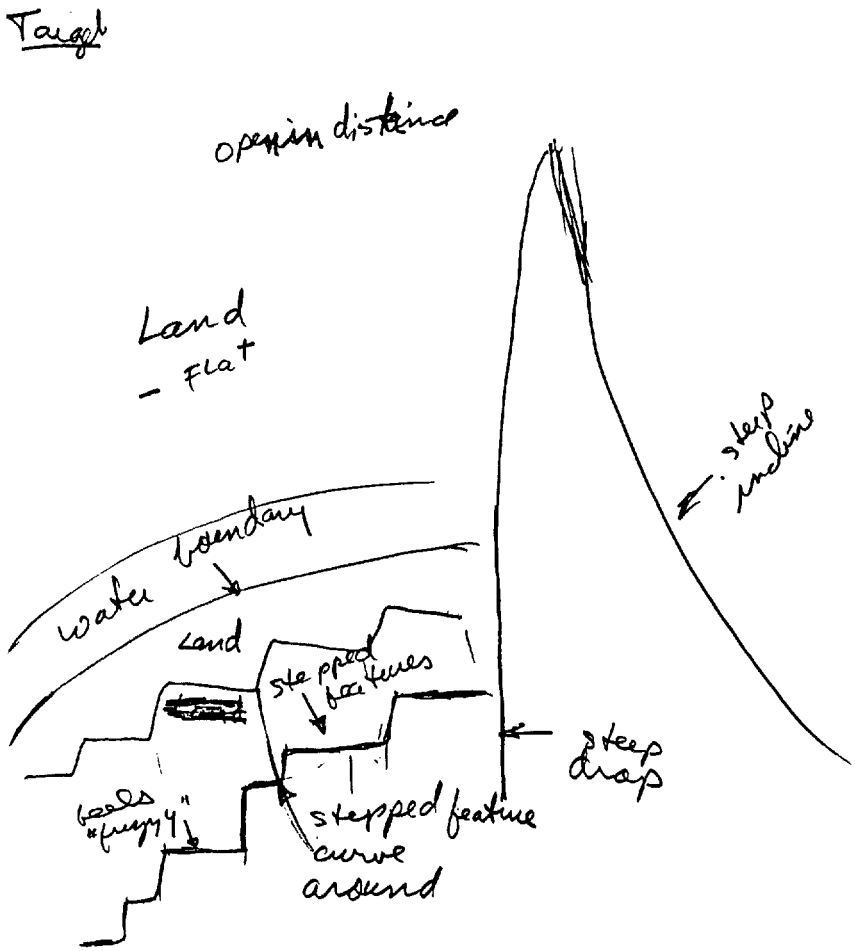


Figure B10. Page five of response (Session 9005, Target 1005).

TABLE B2
TARGET–RESPONSE 9005

| Element | Name | Target | Response | $T_{\alpha=0.2}$ | TΩR |
|---|---|---|---|---|---|
| 14 | Spire, minaret, tower | 0.00 | 0.20 | 0 | 0.00 |
| 20 | Roads | 0.10 | 0.10 | 0 | 0.00 |
| 32 | Urban | 0.80 | 0.70 | 1 | 0.70 |
| 38 | Canal, manmade waterway | 0.00 | 0.10 | 0 | 0.00 |
| 44 | Town, village | 0.00 | 0.30 | 0 | 0.00 |
| 45 | City | 0.90 | 0.70 | 1 | 0.70 |
| 46 | Single peak | 0.00 | 0.20 | 0 | 0.00 |
| 47 | Hills, slopes, bumps, mounds | 0.00 | 0.10 | 0 | 0.00 |
| 54 | Unbounded large expanse water | 0.00 | 0.40 | 0 | 0.00 |
| 56 | Partially bounded water | 0.30 | 0.30 | 1 | 0.30 |
| 58 | River, stream, creek | 0.00 | 0.40 | 0 | 0.00 |
| 59 | Coastline | 0.00 | 0.20 | 0 | 0.00 |
| 60 | Vegetation, trees | 0.20 | 0.20 | 1 | 0.20 |
| 64 | Blue | 0.25 | 0.00 | 1 | 0.00 |
| 65 | Green | 0.20 | 0.00 | 1 | 0.00 |
| 67 | Brown, beige | 0.50 | 0.00 | 1 | 0.00 |
| 69 | White | 0.10 | 0.00 | 0 | 0.00 |
| 70 | Grey | 0.10 | 0.00 | 0 | 0.00 |
| 80 | Smooth | 0.10 | 0.00 | 0 | 0.00 |
| 81 | Fuzzy | 0.00 | 1.00 | 0 | 0.00 |
| 82 | Grainy, sandy, crumbly | 0.00 | 1.00 | 0 | 0.00 |
| 83 | Rocky, ragged, rubbled, rough | 0.00 | 1.00 | 0 | 0.00 |
| 91 | Congested, cluttered, busy | 0.70 | 0.70 | 1 | 0.70 |
| 94 | Open, spacious, vast | 0.10 | 1.00 | 0 | 0.00 |
| 95 | Ordered, aligned | 0.00 | 0.30 | 0 | 0.00 |
| 96 | Disordered, jumbled, unaligned | 0.30 | 0.00 | 1 | 0.00 |
| 97 | Buildings, structures | 0.80 | 0.90 | 1 | 0.90 |
| 98 | Rise, vertical rise, slope | 0.00 | 1.00 | 0 | 0.00 |
| 99 | Flat | 0.50 | 1.00 | 1 | 1.00 |
| 100 | Light/dark areas | 0.10 | 0.00 | 0 | 0.00 |
| 101 | Boundaries | 0.20 | 1.00 | 1 | 1.00 |
| 102 | Land/water interface | 0.30 | 1.00 | 1 | 1.00 |
| 103 | Land/sky interface | 0.10 | 0.10 | 0 | 0.00 |
| 104 | Single predominant feature | 0.10 | 0.40 | 0 | 0.00 |
| 106 | Manmade, altered | 0.80 | 0.80 | 1 | 0.80 |
| 107 | Natural | 0.20 | 0.20 | 1 | 0.20 |
| 108 | Rectangle, square, box | 0.70 | 1.00 | 1 | 1.00 |
| 111 | Cross-hatch, grid | 0.30 | 0.00 | 1 | 0.00 |
| 112 | Circle, oval, sphere | 0.10 | 0.00 | 0 | 0.00 |
| 116 | Semicircle, dome, hemisphere | 0.10 | 0.30 | 0 | 0.00 |
| 118 | Repeat motif | 0.40 | 0.80 | 1 | 0.80 |
| 119 | Stepped | 0.20 | 1.00 | 1 | 1.00 |
| 120 | Parallel lines | 0.30 | 0.30 | 1 | 0.30 |
| 121 | Vertical lines | 0.50 | 1.00 | 1 | 1.00 |
| 122 | Horizontal lines | 0.10 | 0.00 | 0 | 0.00 |
| 123 | Diagonal lines | 0.10 | 0.20 | 0 | 0.00 |
| 125 | Inverted V-shape | 0.00 | 0.20 | 0 | 0.00 |
| 127 | Arc, curve | 0.30 | 1.00 | 1 | 1.00 |
| 128 | Wave form | 0.00 | 0.10 | 0 | 0.00 |
| Totals | | | 21.20 | 22.00 | 12.60 |

Accuracy = 0.573
Reliability = 0.594
Figure of merit = 0.340

## APPENDIX C
## "GROUND TRUTH" INSTRUCTION AND CODING FORM

*Analysts' Instructions for Remote-Viewing Series 900X*

Thank you for helping us perform a *post hoc* assessment of a series of remote viewings. The targets were actually 35-mm slides that were attached to a photomultiplier, a device to measure small amounts of light. We were searching for possible physical correlates to remote viewing.

You will find in your packet 6 remote viewing responses labeled 9001–9006 respectively. Also shown is the target number of the intended photograph. We have supplied the original, rather than the 35-mm slide.

We would like you to make a *subjective* judgment as to the degree of correspondence between the remote viewing response and its associated target. Familiarize yourself with the task by first looking at all the responses and their intended targets. Then, on a session-by-session basis, rate your assessments. You are completely free to define what is meant by "Degree of Correspondence." Indicate your judgment by marking one line across the appropriate continuous scale shown below. A vertical line near the "None" end of the scale will indicate that you feel there is very little correspondence between that response-target pair. Likewise a vertical line near the "Complete" end of the scale will indicate that you feel that there is a significant degree of correspondence.

Many of the responses begin with a little information and build toward a composite drawing at the end. Please assess the response in its entirety as best you can. Thank you again.

| SESSION | DEGREE OF CORRESPONDENCE | TARGET |
|---|---|---|
| | None — Complete | |
| 9001 | ⊢——————⊣ | 1034 |
| 9002 | ⊢——————⊣ | 1042 |
| 9003 | ⊢——————⊣ | 1065 |
| 9004 | ⊢——————⊣ | 1094 |
| 9005 | ⊢——————⊣ | 1005 |
| 9006 | ⊢——————⊣ | 1024 |

## References


DAWES, R. M. (1988). *Rational choice in an uncertain world.* New York: Harcourt, Brace, Jovanovich.

HONORTON, C. (1975). Objective determination of information rate in psi tasks with pictorial stimuli. *Journal of the American Society for Psychical Research,* **69,** 353–359.

HUBBARD, G. S., MAY, E. C., & FRIVOLD, T. J. (1987). Possible photon production during a remote viewing task: A replication experiment. Final Report, SRI Project 1291, SRI International, Menlo Park, California.

HUMPHREY, B. S., MAY, E. C., TRASK, V. V., & THOMSON, M. J. (1986). Remote viewing evaluation techniques. Final Report, SRI Project 1291, SRI International, Menlo Park, California.

HUMPHREY, B. S., MAY, E. C., & UTTS, J. M. (1988). Fuzzy set technology in the analysis of remote viewing. *Proceedings of the 31st Annual Convention of the Parapsychological Association* (pp. 378–394).

JAHN, R. G., DUNNE, B. J., & JAHN, E. G. (1980). Analytical judging procedure for remote perception experiments. *Journal of Parapsychology,* **44,** 207–231.

MAY, E. C. (1983). A remote viewing evaluation protocol. Final Report (revised), SRI Project 4028, SRI International, Menlo Park, California.

MAY, E. C., HUMPHREY, B. S., & MATHEWS, C. (1985). A figure of merit analysis for free-response material. *Proceedings of the 28th Annual Convention of the Parapsychological Association,* (pp. 343–354).

PUTHOFF, H. E., & TARG, R. (1976). A perceptual channel for information transfer over kilometer distances: Historical perspective and recent research. *Proceedings of the IEEE,* **64**(3), 329–354.

SOLFVIN, G. F., KELLY, E. F., & BURDICK, D. S. (1978). Some new methods of analysis for preferential-ranking data. *Journal of the American Society for Psychical Research,* **72**(2), 93–110.

TARG, R., PUTHOFF, H. E., & MAY, E. C. (1977). State of the art in remote viewing studies at SRI. *1977 Proceedings of the International Conference of Cybernetics and Society* (pp. 519–529).

ZICK, R., CARLSTEIN, E., & BUDESCU, D. V. (1987). Measures of similarity among fuzzy concepts: A comparative analysis. *International Journal of Approximate Reasoning,* **1**(2), 221–242.



*SRI International*
*333 Ravenswood Av.*
*Menlo Park, CA 94025*

*Division of Statistics*
*University of California, Davis*
*Davis, CA 95616*