[Dec 7 1966]

MEMORANDUM

To: Messrs. Condon, Cook, Low, Roach, Scott, Wertheimer

Subject: A framework for the analysis of UFO data

One way in which to conceptualize our problem is that we are dealing with a gigantic data matrix. There may be tens of thousands of rows in this matrix, no matter whether we regard the proper row-entry as a report (by one person) or as an event (possibly with multiple witnesses). In practice we will be dealing with a set of reports, and will have to face the problem of partitioning them among a smaller number of events. If our primary classifications are chronological and geographical, and we insist that these items of information not be missing, reports that stem from the same event will stand fairly close together in any list. I would assume, however, that such lists would be computer-prepared. For office filing purposes we should simply assign sequential "accession numbers," which will be unique and invariant. I will return below to the question of manipulating the rows of this data matrix.

The columns of this data matrix represent various facets of an event (or report) that might be separately isolated and characterized. (I do not mean to imply either that these are mutually exclusive or jointly exhaustive of the content of UFO reports.) There seem to be, however, three logically distinct kinds of information that will be tapped by the columns of the matrix:

1) Situational information.
2) Observational information.
3) Explanatory information.

These may be defined adequately for the present purpose by reference to the attached lists of variables that would be classifiable under each of these headings A second task we face is that of making these lists relatively complete, which may well lead us to upwards of a hundred columns of data. Even so, we must remain alert to the fact that we are missing columns as well as rows from a theoretically complete matrix.

The entries in the cells of this data matrix will not always be numbers. Most of the cells will be empty. Most of the visible entries, especially for the observational variables, will be in verbal and even narrative form. (As we learn how to quantify this material we will simply define new variables to accomodate the numbers.) Even at best, this does not provide a matrix amenable to most conventional matrix algebra but then -- we are volunteers! The point is that this matrix does provide a home for each piece of data, and is immediately amenable to certain interesting and computer-programable operations. A further point is that this data matrix is limited to relatively factual kinds of information. By the application of various operations to this data matrix -- operations that they will be expected to define -- various investigators may be expected to attain various interpretations of the UFO phenomenon.

In our discussions, Vallee has emphasized and others have supported the value to be derived from a search for correlations between different columns within the first two-thirds of the matrix. In generating such correlations, the rows should typically be weighted on the basis of information in other columns, including the "explanatory." I.e., events should be

weighted inversely to the total plausibility of "conventional" explanations.
It is within the capability of this approach to look for "waves" of sight-
ings, to examine the periodicity of any waves, to assess the contribution
of positive feedback via social circuits to wave phenomena, to search for
typical recurrent patterns of observations either by themselves or in com-
bination with observer characteristics, to compare-contrast the files of
Blue Book, NICAP, APRO, et al., to predict the time-space coordinates of
future events (undoubtedly at a low level of certainty), etc, etc. The major
concern with this approach is that it requires a large sample of data rows,
because the posited relationships are not simple, and each row must be suf-
ficiently well-evaluated to determine its appropriate weight. We are unlike-
ly to complete any study of this kind in time for it to contribute to the
planning of other studies to be completed by January 1968. However, I sus-
pect some of the results of this approach are worth getting for their own
sake.

Obviously a more fundamental operation is the one that can be carried
out within any given row, with the idea of filling in the explanatory cells.
Again, Vallee's 25 questions provide a programmable first approximation, but
this will often be ineffective because of missing information. In the case
of some of the more imaginative specifications for IFO's, skilled analytical
efforts and additional field investigations may be required before a low
probability can be rationally adduced. It does not seem justifiable to me
that this intensity of high-level effort should be expended on more than a
selected sample of events. The primary means of acquiring an adequate sample
size for correlational studies should be routine clerical and computer pro-
cessing of larger numbers of raw reports, which seem to be very easy to come
by.

The other approach to our data matrix, emphasizing rows instead of col-
umns, is advocated implicitly by most authors of UFO books and has been most
clearly represented in our discussions with NICAP. Using Hynek's termin-
ology, however, the object of this game is to accumulate a sufficient sample
of reliably-reported, carefully-investigated, "strange" events to over-tax
our credulity in the sufficiency of available IFO explanations. Entirely
apart from the question of the absolute "amount of evidence" required for
this outcome, we must be prepared to tolerate probably individual differences
among ourselves in our judgments of this quantity, which depends on tempera-
ment as well as training. Of course, as scientists, it is incumbent upon us
to pursue every line of investigation or reasoning that could alter our
probability estimate of an IFO explanation in either direction; this is the
only sense in which we can claim to be impartial and objective. We are ob-
viously committed to play this game; the only questions have to do with
strategy.

Let me suggest here that there are simple indices available from our
data matrix to measure the reliability, complexity, and strangeness of events
that have been recorded in it by clerical procedures, so that a computer
might at any stage of our work produce a listing of events ranked according
to their apparent potential contribution to the objective. Such a ranking
would necessarily be imperfect, but would be reasonably trustworthy in terms
of the relatively few events floated to the top of the list. We would then
be in a position to assess these as a group in terms of the kind and amount
of effort likely required finally to confirm or deny IFO status, and to al-

locate or select our resources to best advantage. New events would compete for our attention on the same basis as old ones, giving due recognition to their natural advantages. Likewise with events in any other competing categories.

Until such time as the credulity of some of us may become over-taxed, this should be our primary mode of operation. If UFO's are not real, it should be impossible to go further. If they are real, we should reach the next stage at the earliest possible time commensurate with our level of support and our own temperaments. If and when we have succeeded in isolating a "convincing" series of unidentifiable UFO's, we must face the additional task of dreaming-up engineering specifications for "possible" UFO's. "Possible" in this context simply means "in violation of a minimum number of conventionally-assumed engineering limitations." We may then proceed to see how many UFO's would become IFO's, and to devise instrumentation that might distinguish one model of "possible" UFO from another. It would be in this context that results based on columnar analysis of our data matrix would begin to be useful.

In view of what I take to be the political realities of our situation, I feel we should set our priorities so as to maximize the possibility of reaching this advanced stage within the life of our contract. I do not wish by this statement either to prejudice you or to expose my own possible bias. The point is that, if we conclude UFO's are not real, we will not ask for more money and the matter will rest there for some time. Ultimately, the most costly risk would be for us to reach this conclusion when or if UFO's are real. Therefore we should conduct ourselves, not as the AIR Force, but so as to be as sure as possible that we recognize any genuine reality to the existence of UFO's

All of which leads me to the following specific recommendations:

1) At least one of us (obviously I'm volunteering) should accept the responsibility, with whatever help a computer can provide, of keeping us posted on our progress and opportunities, so that we may consciously pursue the above-outlined strategy.

2) To begin with, our data matrix should contain the events already nominated to us by Hynek, Vallee, Keyhoe, Hall, and others. (N < 100). At this point, I don't know which of these events should command our attention. In any case, our need for experience as investigators is an immediate major consideration.

3) The data matrix should immediately be supplemented by the events reported through various authors, and as recorded by them. This material has already been at least partially organized by Olsen. (N ≈ 1000). By this time our coding scheme (column-definitions) should be pretty well-established.

4) The results of our own investigations and deductions would be constantly fed in.

5) Without detracting from our investigative efforts, we should sample various further sources for the enlargement of the data matrix -- Blue Book, NICAP, AFRO, etc -- to establish priorities for more complete sampling.

6) These files would then be incorporated in the priority indicated.

7) The ultimate source of new reports would be direct public solicitation of old as well as current sightings. (Data collected in this way

would be of particular value because of the de-synchronized feedback.)

8) New columns would be added to the matrix at any time for any good reason.

9) Periodically, the material would be searched to list the most promising events for further immediate consideration.

10) When the file became adequate, column-oriented analyses would be undertaken. Programs and pilot runs of this kind would be worked out in advance of this point, so as to be run-able for the record just in time for our final report.

If, having read this, you feel I've taken four pages merely to belabor the obvious, please don't hesitate to say so.

December 7, 1966   D. R. Saunders