

Survey shows general agreement on the meaning of "probable" and some equivalents, elsewhere much disagreement.

THE DEFINITION OF SOME ESTIMATIVE EXPRESSIONS

David L. Wark

Finished intelligence, particularly in making estimative statements, uses a number of modifiers like "highly probable," "unlikely," "possible" that can be thought of as expressing a range of odds or a mathematical probability, and these are supplemented by various other expressions, especially verb forms, conveying the sense of probability less directly—"may," "could," "we believe." Certain other words express not probability but quantity, imprecisely but perhaps within definable ranges—"few," "several," "considerable." Some people object to any effort to define the odds or quantities meant by such words. They argue that context always modifies the meaning of words and, more broadly, that rigid definitions deprive language of the freedom to adapt to changing needs.

It is possible, however, to state the definitions in quantitative terms without making them artificially precise. And if two-thirds of the users and readers of the word *probably*, for example, feel it conveys a range of odds between 6 and 8 out of 10, then it is more useful to give it this definition than to define it more or less tautologically in terms of other words of probability. This would not deny to context its proper role as the arbiter of value, but only limit the range of its influence. Nor would it freeze the language in perpetuity; as the meanings of the words evolved the quantitative ranges could be changed.

This article describes the results of a survey undertaken to determine if such words are indeed understood as measurable quantities and if so to ascertain the extent to which there is a consensus about the quantitative range of each. A three-part questionnaire on the subject was distributed in the intelligence community—to INR/State, the DIA Office of Estimates, and five CIA offices—and a simplified version of it was sent to policy staffs in the White House, State, and

the Pentagon. Responses were received from 240 intelligence analysts and 63 policy officers.

The responses showed a satisfactory consensus with respect to various usages of *likely* and *probable*, phrases expressing greater certainty than these, and modifications of *chance*—good, better-than-even, slight. There was no satisfactory agreement on the meaning of *possible* or a wide variety of verb forms such as *we believe* and *might*. There was also little agreement on the non-odds quantitative words such as *few* and *many*. The policy offices consistently assigned lower probabilities than intelligence analysts did. Correlation between values assigned in and out of context was good.

The Questionnaire

Part One of the questionnaire listed 41 expressions that might be thought of as indicating odds and offered the choice of 0, 10, 20, etc. through 100 as the percentage probability or chances out of 100 signified by each. If the respondent believed that no quantitative answer was satisfactory he could mark "Not Applicable" instead. These expressions of course had to be judged without benefit of context, but in order to check on the validity of such judgments some of them were repeated in Part Two, where they were included in 17 sentences taken from intelligence documents which had been produced in six different offices of the community. The names of all persons and countries in the sentences were changed to sterilize them against bias. Part Three then listed nine expressions of magnitude not referring to probability and offered an assortment of ranges for each.

The idea of a consensus is relative, but for purposes of Parts One and Two it was defined as requiring 70% or more of respondents to name odds within 10 points, plus or minus, of the most frequent response. If the odds or chances most frequently specified for *possibly* were 50 out of a hundred (as they were) and 70% of all the responses had fallen within the range 40 to 60, the requirements for a consensus on this word would have been satisfied. Only one figure was recorded for each question: when an answer was ranged by marking several adjacent figures, it was recorded as the mean. Mr. Kent's range of 10 to 90 for *possible* would thus have been recorded as 50. Definitions were also considered invalidated by 20% or more of "Not Applicable" responses rejecting the question.

The replies were tabulated in four categories in descending order of valid definition, as follows:

Category A—a consensus including 90% or more of all respondents.

Category B—a consensus including 70% to 89% of all respondents.

Category C—no consensus, but fewer than 20% of respondents marked "Not Applicable."

Category D—no consensus, and 20% or more of respondents marked "Not Applicable."

Findings

The following tables summarize the findings of the survey. After each expression from Parts One and Two are shown the odds most frequently specified and the percentage of respondents within 10 points of that. For questions submitted to policy officers as well as analysts, their responses are shown separately. The expressions of magnitude in Part Three are listed with the percentage of "Not Applicable" responses and the most frequent response for each.

Of the 41 expressions in Part One three fell into Category A (super-consensus), thirteen into Category B (consensus), seventeen into Category C (no consensus), and eight into Category D (rejected as indefinable). From Part Two five expressions in context fell into Category B, twelve into Category C, and three into Category D. All the quantitative phrases in Part Three were rejected as not measurable by 20% or more of the respondents except for *next few years* and *next year or so*. Though rejected by only 7%, *next few years* found no consensus: 19% marked 2 to 3 years, 30% 2 to 4 years, and 34% 2 to 5 years. *Next year or so* meant 1 to 2 years to two-thirds of the respondents, 1 to 3 years to the rest.

PART ONE (No Context)

EXPRESSION	ODDS—MOST FRE- QUENT RESPONSE		PERCENT AGREEING WITHIN 10 POINTS	
	Analyst	Policy	Analyst	Policy
Category A (90%–100% Consensus)				
Almost Certainly	90	90	99%	94%
Are	100	100	96%	92%
Will	100	100	91%	91%
Category B (70%–89% Consensus)				
Probably	75	70	90%	86%
Probably not	20	20	85%	76%
Probably will	80	—	85%	—

EXPRESSION	ODDS—MOST FRE- QUENT RESPONSE		PERCENT AGREEING WITHIN 10 POINTS	
	Analyst	Policy	Analyst	Policy
Category B (70%–89% Consensus) (Continued)				
Highly probable	90	85	83%	87%
Likely	70	—	83%	—
Undoubtedly	100	90	81%	86%
Good chance	70	70	81%	81%
Highly likely	90	80	80%	81%
Unlikely	20	20	80%	79%
Seems likely	70	—	80%	—
Better than even chance	60	60	78%	87%
Some slight chance	10	10	77%	79%
May	50	—	73%	—
Category C (No Consensus)				
Seems unlikely	20	—	68%	—
Might	50	50	66%	59%
May indicate	50	—	66%	—
Could be expected	60	—	65%	—
Expect	80	—	64%	—
Could	50	50	60%	56%
Must	80	—	59%	—
Evidently	70	—	59%	—
Apparently	70	—	58%	—
Suggests	60	—	58%	—
Believe	70	70	55%	54%
Should	70	—	54%	—
Possibly	50	50	53%	51%
Might be expected	50*	—	51%	—
Indicates that	70	—	51%	—
Might be anticipated	50	50	56%	50%
Apparently is intent	60*	—	50%	—
Serious possibility	60*	70	49%	55%
Category D (Rejected)				
Estimate	75	70	56%	57%
Seems	50	—	55%	—
Ought	60*	—	41%	—
Feel	50†	—	35%	—
Reportedly	50	50	35%	52%
Somewhat	50*	—	27%	—
Ostensibly	50	—	20%	—

* The most frequent response to these questions was of the same order as a second most frequent. Graphed, the responses would show two frequency peaks, forming, in statistical terminology, a "bimodal" curve.

† The curve for this response is trimodal.

PART TWO (In Context)

EXPRESSION (IN CONDENSED CONTEXT)	ODDS—MOST FRE- QUENT RESPONSE		PERCENT AGREEING WITHIN 10 POINTS	
	Analyst	Policy	Analyst	Policy
Category B (70-89% Consensus)				
We believe the chances are good that . . .	70	—	86%	—
We believe . . . will not be . . .	80	80	76%	63%
Undoubtedly, . . . will not be . . .	100	—	76%	—
We estimate . . . will not be . . .	80	70	74%	70%
Barring . . . the economy will probably continue . . .	80	—	71%	—
Category C (No Consensus)				
Apparently, . . . will not be . . .	70	—	68%	—
If . . . continue . . . , the president might . . . be willing . . .	50	50	65%	54%
. . . might also take . . . action . . .	50	—	62%	—
. . . references . . . to undiminished im- portance . . . suggest a belief . . .	60*	—	59%	—
It is possible that . . . will become . . .	50	50	56%	57%
. . . visit . . . indicates that . . . is be- ing . . . ¹	70	—	53%	—
. . . visit suggests . . . progress . . . ¹	60	—	51%	—
We believe . . . there is a possibility that . . .	50	50	50%	43%
. . . speech . . . conveyed the impression that . . . ²	60*	—	46%	—
. . . comments suggest . . . changes may well be less than speech . . . might indi- cate . . . ³	70*	65	43%	40%
Category D (Rejected)				
. . . comments suggest . . . that . . . gov- ernment is not committed . . . ⁴	0†	50†	18%	25%

¹ The full context on these questions was the sentence, "Although lacking the drama of visits by top leaders, the travel of these delegations to Albania indicates that the momentum of the Albanian-Polish rapprochement is being maintained and suggests that some progress is being made in reducing the area of remaining ideological differences." Respondents were asked to specify the probability that Albania and Poland were headed toward a rapprochement and the probability that the ideological differences would be settled.

² Respondents were asked for the probability that the speaker believed what he conveyed.

³ Respondents were asked for the probability that changes would be minor.

⁴ Respondents were asked for the probability of that to which the "government is not committed." The full context is given on page 73.

* Bimodal.

† Trimodal.

EXPRESSION (IN CONDENSED CONTEXT)	ODDS—MOST FRE- QUENT RESPONSE		PERCENT AGREEING WITHIN 10 POINTS	
	Analyst	Policy	Analyst	Policy
Category D (Rejected) (Continued)				
This raises the question whether . . . they might . . .	50	—	51%	—
We do not expect them to change . . . [‡]	90†	—	22%	—
Cuba has allegedly bought . . .	50	—	38%	—

[‡]This question was a non-sequitur. The full context is given on page 73.
† Trimodal.

PART THREE (Words of Magnitude)

EXPRESSION	PERCENT REJECTING	MOST FRE- QUENT RESPONSE
Considerable	47%	10-100
Many	40%	10-1000
Substantial (portion)	36%	20-50%
Significant (portion)	34%	20-50%
Limited (portion)	30%	2-10%
Several	27%	2-5
Few	28%	2-4
Next few years	7%	2-5 years
Next year or so	1%	1-2 years

The difference between the good consensus on a set of odds for one expression and no consensus on another shows up clearly when the odds are graphed according to how frequently each set was specified in the responses to a question. When 70% of all responses fall within 10 points of the most frequent one, the graph has a steep curve and a narrow base. The high, narrow peak indicates a clearly defined consensus, whereas a broad-based curve with a single peak shows less agreement and a curve with several peaks reflects clear differences about what the word means.

Steady retrogression from consensus can be seen in graphs of sample responses from successive categories. Following are these seven from Parts One and Two.

	CATEGORY
Out of context:	
Almost Certainly	A
Probably	B
Possibly	C
Serious Possibility	C
Seems	D

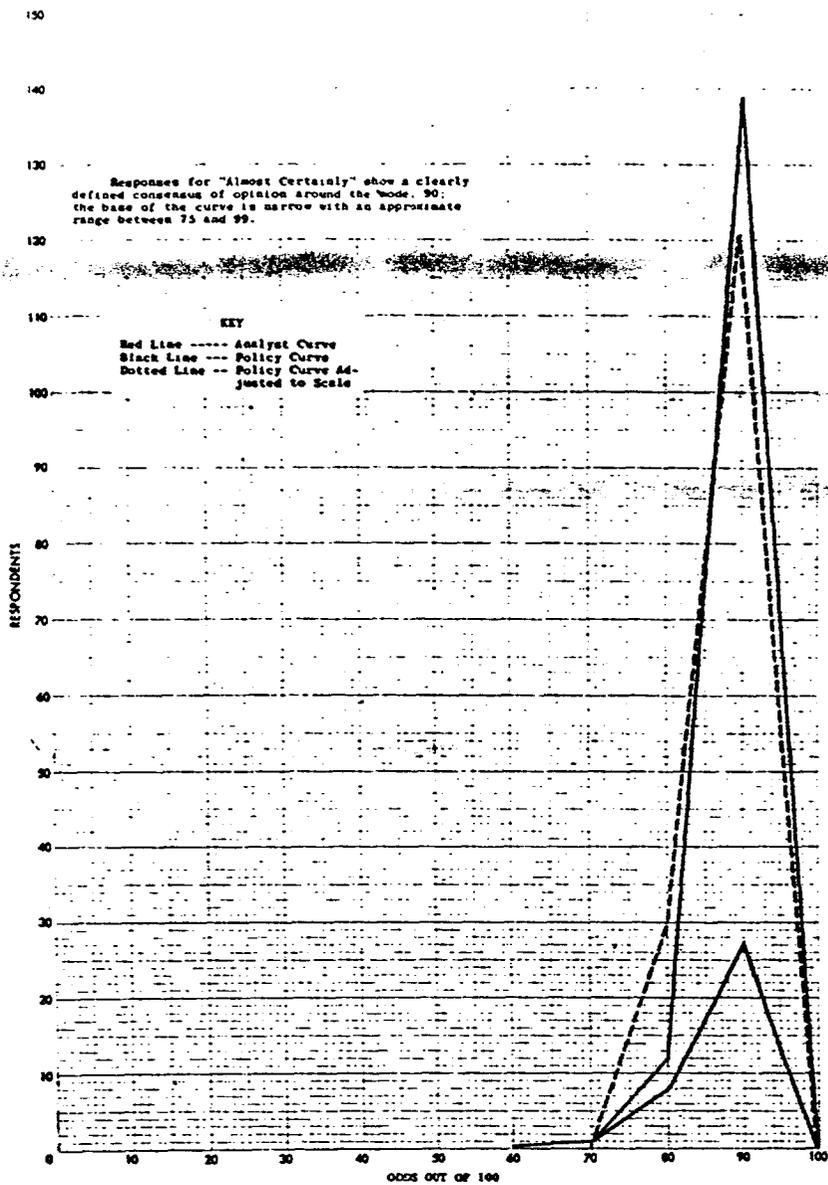
CATEGORY

In context:

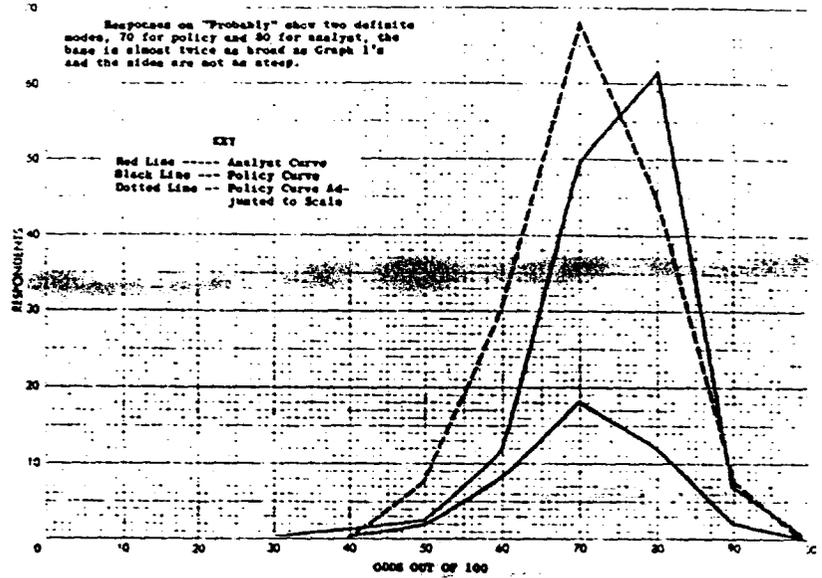
"The North Koreans have thus far shown marked respect for US power, and we *do not expect* them to change this basic attitude" expresses what probability that the North Koreans will continue provocations against South Korea? D

"At the same time, the reservations conveyed in the military comment suggest that the practical military changes resulting from the new line may well be less dramatic than the tone of de Gaulle's speech might indicate—and that in any event, his government is not committing itself to a one-weapon system of defense" expresses what probability that the military will have a one-weapon system? D

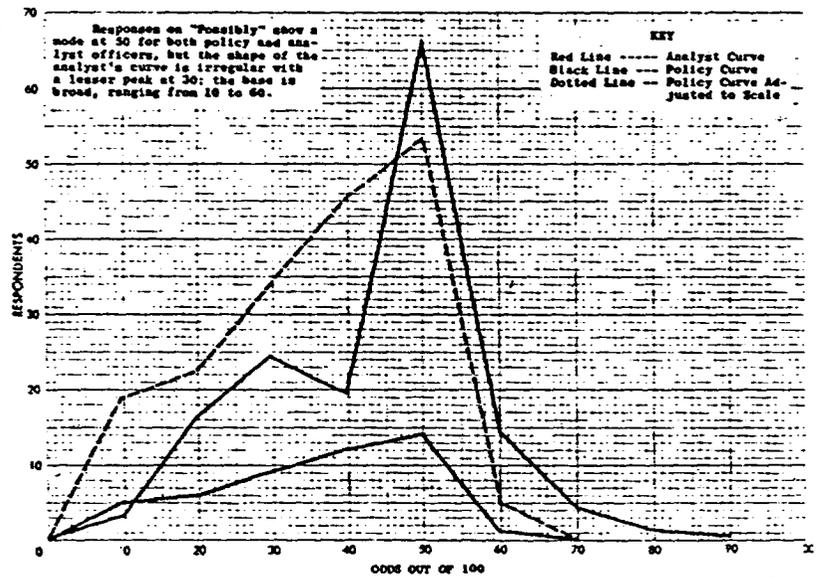
The red line in each graph traces the response pattern of 239 analysts, the black line in the first four that of 63 policy officers. The dotted black line is the latter adjusted to scale. "Mode" designates the peaks of most frequent response.



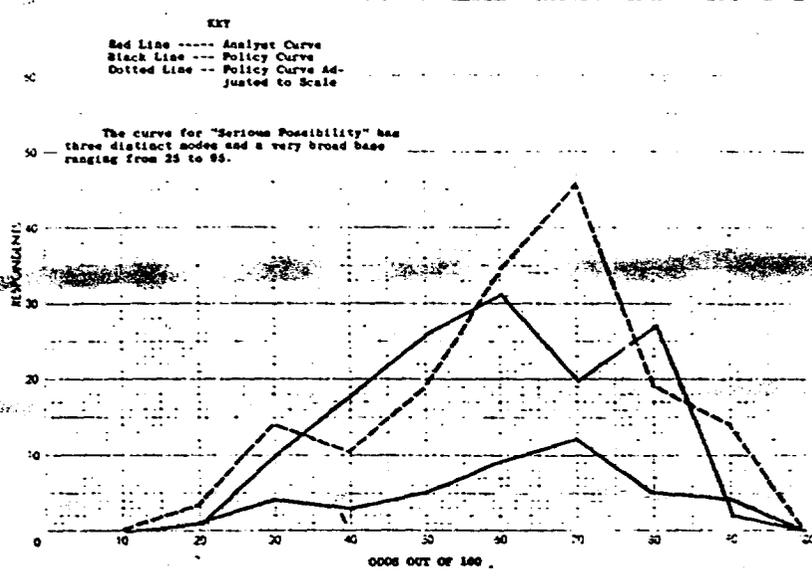
GRAPH No. 1. Category A: *Almost Certainly* (Significant Range 75-99).



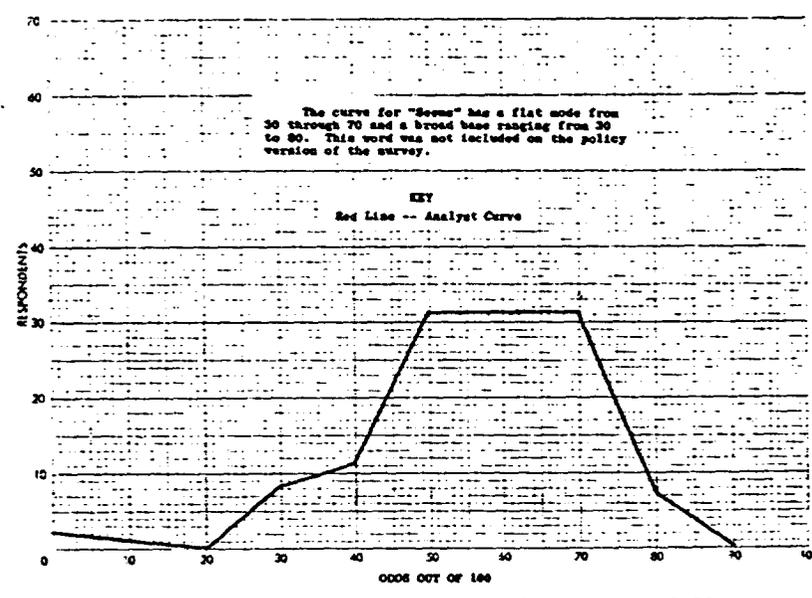
GRAPH No. 2. Category B: *Probably* (Significant Range 50-90).



GRAPH No. 3. Category C: *Possibly* (Significant Range 10-60).

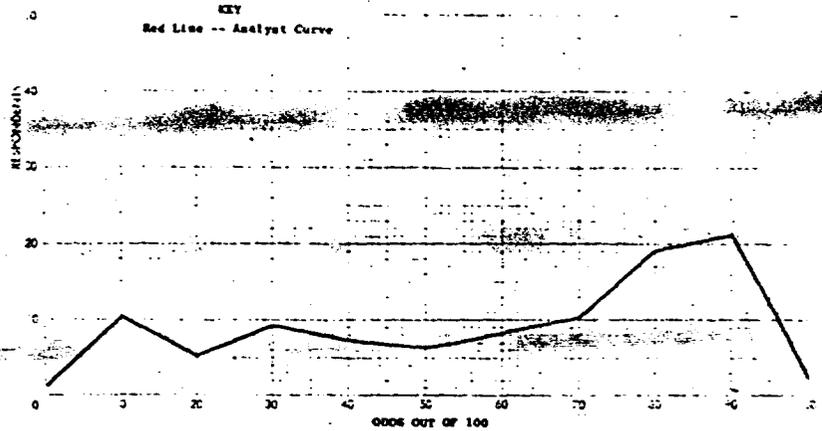


GRAPH No. 4. Category C: Serious Possibility (Significant Range 25-95).



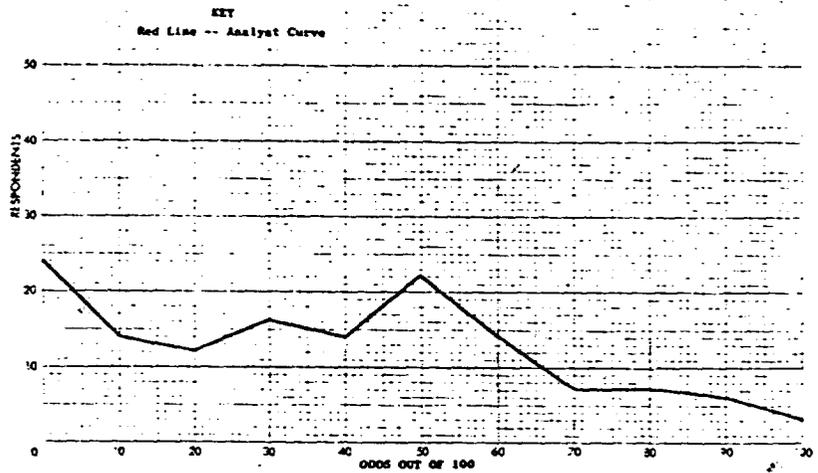
GRAPH No. 5. Category D: Seems (Significant Range 30-80).

This question was included on the analyst version only. The curve ranges from 0 to 100 and shows no clearly defined mode. There is no consensus for the statement of probability.



GRAPH No. 6. Category D: Korean Question (Significant Range 5-95).

On this complicated question the analyst curve ranges from 0 to 100 and shows no clearly defined mode. There is no consensus for the statement of probability.



GRAPH No. 7. Category D: Question with Suggest (Significant Range 0-90).

Conclusions

Of the 303 questionnaires returned, only one indicated that no quantitative equivalent was suitable for any of the probabilistic expressions. All others selected sets of odds for at least half of those listed in Part One, and 80% did so for two-thirds of them. Even though a number who disapprove of quantitative definitions probably just did not bother to return their questionnaires, the results appear to indicate that the vast majority in the intelligence community consider it legitimate to think of such expressions in quantitative terms.

On the other hand, although more than 70% of both analysts and policy officers agreed within a 20-point range on the expressions in Categories A and B, the results for some offices on the analytical side did not agree with the consensus for all analysts, and there were similar exceptions among the policy offices. So when an analyst in one office uses the word *probably*, policy officers and analysts in other offices do not necessarily interpret the word to mean the same thing. In Categories A and B, however, the differences are usually not great. There follows the quantitative definition—most frequent plus and minus 10—of expressions on which there was found to be a satisfactory consensus.

	<u>CHANCES OUT OF 100</u>
Are	90-100
Will	90-100
Almost Certainly	80-100
Undoubtedly	80-100
Highly Likely	75-95
Highly Probable	75-95
Probably Will	70-90
Probably	60-80
Likely	60-80
Good Chance	60-80
Seems Likely	60-80
Better Than Even Chance	50-70
May	40-60
Probably Not	10-30
Unlikely	10-30
Some Slight Chance	0-20

The out-of-context definitions in Part One were spot-checked by the sentence questions of Part Two. The results are not conclusive: only one sentence was provided for context, and there was no way of telling if respondents were influenced by personal knowledge of the subject matter. But despite these limitations, because the most

frequent definitions in and out of context agreed within 10 points, it appears that nearly the same meanings were conveyed either way. The comparison appears below.

	MOST FREQUENT RESPONSE				
	IN CONTEXT		ALONE		
	Analyst	Policy	Analyst	Policy	
Undoubtedly	100	—	Undoubtedly	100	90
Believe	80	80	Believe	70	70
Estimate	80	80	Estimate	75	70
Apparently	70	—	Apparently	70	—
Indicates that	70	—	Indicates that	70	—
Believe the chances are good	70	—	Good chance	70	70
Possible	50	50	Possibly	50	50
Might	50	—	Might	50	50

Although the coupling of a verb of opinion with an expression of odds, as in "We believe the chances are good," seemed not to affect the meaning of the latter for the respondents to the survey questions, this writer agrees with Mr. Kent's purist that the doubling up of probabilistic words is potentially confusing and should be avoided. The response pattern on the Korean question (page 73 and Graph 6) has an interesting side light in that the probability queried does not follow from the estimative sentence. The questionnaire was not designed to test the propensity of analysts and policy officers to draw unsubstantiated conclusions, but in this one instance only 35% of the respondents showed they recognized the non-sequitur by marking "Not Applicable."

The survey showed that for expressions on which there was a consensus (and some others) the most frequent response was the same from policy officers as from analysts within 10 points plus or minus. But where differences did occur, the policy definitions were consistently on the conservative side; see the following examples.

	MOST FREQUENT RESPONSE	
	Analyst	Policy
Undoubtedly	100	90
Highly Probable	90	85
Highly Likely	90	80
Probably	75	70
Estimate	75	70

The results from Part Three showed there is little consensus on the common expressions of vague magnitude, at least without the guidance of context.

~~CONFIDENTIAL~~

Estimative Expressions

Faults in the Survey

An effort was made to keep the questionnaire as simple to understand and as short as possible. In Parts One and Three the effort was generally successful, but Part Two was neither simple nor short. Most of the questions in the latter related to specific people and places, and there was danger that respondents would permit their opinions and knowledge of the subject to influence their answers. In addition, several of the estimative sentences were long and involved, carrying the hazard of confusion about what they meant and what was wanted in evaluation of them.

For pragmatic reasons, administration of the survey had to be informal. It is possible that such things as attitudes of supervisors, office collusion, or misunderstanding of the purpose of the survey could have introduced bias. A careful perusal of each of the questionnaires failed to turn up any obvious evidence that such factors influenced the findings. But if it were done again the questionnaire should be modified in Part Two and the conditions under which it is filled out should be controlled and standardized.

~~CONFIDENTIAL~~