The effect of multiple responses and certainty estimates on the integration of visual information[1]

JOSEPH HALPERN[2] AND Z. JOSEPH ULEHLA

UNIVERSITY OF DENVER

Integration of information was investigated in two experiments that employed a multiple-observation, tilt-discrimination task. The integration model of signal-detection theory generally overestimated the improvement in discrimination resulting from additional observations. Further, probabilistic multiple responses appeared detrimental to discrimination. These results differ from earlier findings using different discrimination tasks. Reasons for the differences were discussed in terms of within-trial redundance and interference caused by complex response modes.

Research concerned with integration of information, where a multiple-stimulus observation methodology has been used, suggests that a probabilistic mode of response results in more efficient use of information than does a simple "yes-no" response. Further, this probabilistic mode of response appears to provide data consistent with predictions of the integration model of the theory of signal detection (TSD) (Green & Swets, 1966). Swets and Birdsall (1966) employed an auditory-detection task where each trial contained either six signal-plus-noise time intervals or six noise-alone time intervals. In one treatment group, where a probabilistic mode of response was used, S expressed his decisions in the form of probability ratings. In a second treatment group, the task was identical except that the S made a simple yes-no response. The improvement in discrimination with additional stimulus observations approximated theoretical predictions only for the condition that included a probabilistic response. Discrimination improvement fell short of theory in the condition involving merely a yes-no response. Similar results were obtained by Ulehla, Canges, and Wackwitz (1967) in a conceptual-discrimination task. In the Ulehla et al study, improvement in discrimination fell short of theory for conditions involving a single binary response per trial, but provided a good fit to predictions of the integration model when S made a complex probabilistic response after each of three stimulus presentations per trial. This complex response included both a discrimination judgment followed by a certainty estimate, i.e., a numerical

assessment by S of the likelihood that his discrimination judgment was correct along a continuum ranging from 50% to 100% in steps of 10%.

Ulehla, Halpern, and Cerf (1968) tested the fit of the integration model to a visual-discrimination task where the discrimination involved the judgment of direction of tilt. Although an analysis of operating characteristics suggested the applicability of TSD to this type of task, the integration model was not supported. In two separate experiments, they found that performance on the second observation was close to prediction while performance beyond this second observation fell considerably short of predicted values. The Ulehla et al (1968) study assumed that certainty estimations and multiple responses (i.e., a response and certainty statement following each observation) would provide optimal conditions for the effective utilization of information provided by multiple-stimulus observations. In view of the evidence given by the Swets and Birdsall (1966) and Ulehla et al (1967) experiments, it seemed logical to assume that such multiple responses and certainty estimates would serve to increase attention and vigilance. The possibility exists, however, that these assumptions may be invalid for visual-discrimination tasks. Consequently, the two experiments given here were designed to determine the effect(s) of multiple responses and certainty estimates on integration of information. A further purpose was to determine whether the pattern of improvement in discrimination with additional stimulus observations obtained by Ulehla et al (1968) could be replicated.

EXPERIMENT 1

Subjects
The S sample included 72 University of Denver undergraduate volunteers, randomly assigned to four groups.

Apparatus
The apparatus and stimulus materials were identical to those used by Ulehla et al (1968). A Kodak Carousel slide projector, fitted with an Alphax shutter, was used to project the stimuli upon a white surface. Each stimulus was a 6½-in. line, tilted ½ deg from the vertical as defined by an enclosing

19-in. square in which the line was centered. The enclosing square was included on each slide to provide Ss with a constant reference in order to reduce the differential effects of possible variation in slide positions. Lines termed "left-tilt" were tilted ½ deg from the vertical in the counterclockwise direction; lines termed "right-tilt" were tilted ½ deg in the clockwise direction.

Procedure
Multiple- and single-response conditions were factorially combined with certainty and no-certainty conditions. The resultant four groups included a condition where Ss made decisions concerning direction of tilt after each stimulus observation with certainty estimates accompanying each response, i.e., Ss made a binary prediction and then attached a confidence rating to the prediction; the confidence rating was given along a continuum ranging from 50% (pure guess) to 100% (complete certainty). The decision, as well as the certainty estimate, involved placing pencil marks in appropriate boxes on an answer sheet. The remaining conditions involved multiple responses with no certainty estimates, single responses with certainty, and, finally, single responses with no certainty.

In the multiple-response groups, S responded after each stimulus presentation. In the single-response groups, S responded only after the last stimulus presentation of each trial. In all groups, S was told the correct tilt at the end of each trial. The task was to discriminate the left-tilt lines from the right-tilt lines, and each S received 180 trials. A different slide was exposed for each stimulus presentation, and Ss were told that all five slides of each trial were of the same tilt. The first exposure of each trial followed the last exposure of the preceding trial by 10 sec. Each trial included five 1/100-sec exposures of the stimulus presented at 5-sec intervals.

Results and Discussion
The effect of multiple responses and certainty estimates was evaluated using the index of signal detectability (d'). Within TSD, d' is a basic measure of the discriminability of a pair of alternative stimuli for a given S and is theorized to be independent of response bias. A more complete explanation of d' (and the

Table 1
Values of d' and d' Ratios for Multiple Response Groups

	O_1	O_2		O_3			O_4			O_5		
	d'_1	d'_2	d'_2/d'_1	d'_3	d'_3/d'_1	d'_3/d'_2	d'_4	d'_4/d'_1	d'_4/d'_3	d'_5	d'_5/d'_1	d'_5/d'_4
Multiple Response With Certainty	.81	1.18	1.46	1.25	1.54	1.06	1.42	1.75	1.14	1.66	2.05	1.17
Multiple Response No Certainty	.98	1.37	1.40	1.53	1.56	1.12	1.65	1.68	1.08	1.79	1.83	1.08
Ulehla, et al (1968)	.96	1.36	1.43	1.53	1.64	1.15	1.64	1.74	1.08	1.75	1.91	1.09
TSD Predicts			1.41		1.73	1.23		2.00	1.16		2.24	1.12

integration model) is available in Ulehla et al (1968). Empirical d' values were obtained for each S by referring his hit rates and error rates to tables of d' (Elliott, 1964). Certainty estimates were ignored in the data analysis. Analysis of improvement in discrimination with additional observations focused upon the d' measure for which the integration model furnishes point predictions. Specifically, the model predicts that the d' based on n observations will be $\sqrt{n}$ times the d', based on a single observation. For all analyses, d' values were computed separately for each S and then averaged to yield group means.

The mean d's, computed from the final response on each trial, for the multiple-response groups with and without certainty, were 1.66 and 1.79, respectively. The mean d's for the single-response groups with and without certainty were 1.84 and 2.00, respectively. These results clearly do not support an assumption that multiple responses and certainty estimates enhance discrimination performance. In fact, the sample means reflect superior performance for those groups that made only a single response per trial and/or made no certainty estimates. However, an analysis of variance with d' as the dependent variable did not yield significant F ratios for either of the main effects or for their interaction. Thus, the effects of certainty estimation and multiple response appear to be either nonexistent or, possibly, detrimental. Implications drawn above from multiple-observation experiments involving other types of discrimination were clearly not confirmed for the tilt-discrimination task.

Table 1 gives the mean d' values and d' ratios for the two groups experiencing multiple-response conditions (with and without certainty estimates). The first d' (d'_1) was computed from the judgments of the first stimulus observation of each trial, the second d' (d'_2) was obtained from judgments based upon the first two stimulus observations of each trial, and so on for d'_3 through d'_5. The integration model predicts the d' values for d'_2 through d'_5 to be in the ratio of $\sqrt{2}$ through $\sqrt{5}$, respectively, to d'_1, i.e., $d'_2/d'_1 = \sqrt{2}$ through $d'_5/d'_1 = \sqrt{5}$.

Although there was a monotonic increase in d' with all additional observations, the mean d' improvement for both certainty and no-certainty conditions conformed to theory over two observations only; improvement over subsequent observations was less than predicted. Data obtained in the five-observation experiment of Ulehla et al (1968) are also given in Table 1. The similarity between the d' values and ratios in the present study and those obtained by Ulehla et al (1968) is evident. The similarity lends further support to the pattern of improvement with multiple observations in which integration fulfills predictions only over two observations.

EXPERIMENT 2

Experiment 1 and earlier research using similar tasks have consistently demonstrated that predictions of the integration model are not accurate in describing the effects of stimulus observations beyond the second. Experiment 2 employed conditions where different Ss experienced from one to five observations per trial. The purpose was to explore the possibility that S's knowledge that additional observations will follow modifies the impact of any single observation. An additional purpose involved further exploration of the effects of certainty estimates and multiple responses on integration and discrimination performance.

Method

The basic experimental method was the same as in Experiment 1. The design is perhaps best described by referring to the matrix in Table 2.

The numeral preceding the letter designation indicates the number of stimulus observations. The Ss in the "MC" groups received from one to five observations per trial with a discrimination response and a certainty estimate following each observation. Ss in the "C" groups made a decision following the final observation (with the number of observations varying from one to five) accompanied by a certainty estimate. The Ss in the "M" groups responded after each observation but made no certainty estimates. Finally, Ss in the "R" groups provided only a single discrimination response, which occurred after the final observation for each trial. There were 10 Ss randomly assigned to each of 20 groups; thus, the total sample consisted of 200 undergraduate volunteers randomly assigned to the 20 treatment conditions.

Results and Discussion

The data were generally characterized by substantial between-S variability. There was, however, sufficient consistency to deal with several of the questions to which the experiment was directed.

Values of d' were computed for individuals and then averaged to obtain group means. Table 3 gives the mean d' and d' ratios computed from the final response of each trial for each group. Thus, while Ss, for example, in the 4-MC condition gave a total of four responses, the d' in Table 3 is computed for only this fourth response. The data in each column are, therefore, equated for the number of observations but not for the number of responses.

The d's in Table 3 relate to the effects of multiple responses and certainty estimates on discrimination performance. The table shows that performance was superior in the

Table 2
Experimental Design

	Number of Stimulus Observations per Trial				
	1	2	3	4	5
Multiple Response With Certainty	1MC	2MC	3MC	4MC	5MC
Single Response With Certainty	1C	2C	3C	4C	5C
Multiple Response No Certainty	1M	2M	3M	4M	5M
Single Response No Certainty	1R	2R	3R	4R	5R

Table 3
Mean Values of d' and d' Ratios for the Final Observation for All Groups

Number of Stimulus Observations Per Trial	1 O_1	2 O_2		3 O_3			4 O_4			5 O_5		
	d'_1	d'_2	d'_2/d'_1	d'_3	d'_3/d'_1	d'_3/d'_2	d'_4	d'_4/d'_1	d'_4/d'_3	d'_5	d'_5/d'_1	d'_5/d'_4
Multiple Response With Certainty	1.00	1.06	1.06	1.48	1.48	1.40	1.72	1.72	1.16	1.84	1.84	1.07
Single Response With Certainty	1.06	1.89	1.78	1.61	1.52	.85	2.48	2.33	1.54	1.95	1.84	.79
Multiple Response No Certainty	1.40	1.39	.99	1.59	1.14	1.14	1.89	1.35	1.19	2.04	1.46	1.08
Single Response No Certainty	1.17	1.99	1.70	2.04	1.74	1.03	2.10	1.79	1.03	2.29	1.96	1.09

absence of certainty estimates, i.e., 9 of the 10 groups not giving certainty estimates had higher d's than corresponding "certainty" groups. Further, in the 16 groups where the effects of multiple responses could be assessed, all of the single-response groups yielded higher d's than their multiple-response counterparts. An analysis of variance was performed with d' as the dependent variable and with independent variables defined as "number of stimulus observations" (five levels), "certainty estimates" (two levels), and "multiple responses" (two levels). The analysis provided significant main effects for number of observations [$F(4,180) = 4.19, p < .01$], certainty [$F(1,180) = 4.02, p < .05$], and multiple responses [$F(1,180) = 4.28, p < .05$]. Thus, while additional observations generally aided discrimination performance, the inclusion of certainty estimates and/or multiple responses proved detrimental.

The between-S d' values (across columns) and the d' ratios in Table 3 relate performance to the number of stimulus observations. The ratios are smaller than predicted by theory. This between-S evaluation of the model is, however, somewhat confounded by the large error variance.

The within-S data relevant to an evaluation of the integration model are given in Table 4 where the mean d' for each response for the multiple-response conditions as well as the d' ratios are shown. Here, the data show that the integration model clearly overestimated the improvement in d' that resulted from additional observations.

Several points worthy of discussion may be stressed. It is clear that integration fell far short of the predictions of the model. These results replicate those of Experiment 1 as well as those of Ulehla et al (1968). Further, the results were clearly different from findings of experiments concerned with auditory detection and conceptual discrimination. In these latter situations, complex response conditions provided data that supported the integration model. In the present task, complex response conditions led to inferior discrimination.

There are several potentially important differences between the present "tilt-discrimination task" and the auditory and conceptual detection and discrimination tasks. At a very obvious level, the requirements involved with a complex response tend to interfere with S's intake and storage of information. Specifically, the complex response conditions involved a decision and/or certainty estimate after each observation. This included the placing of check marks in the appropriate places on a prepared form. Another important difference concerns the fact that, within the precision of the apparatus, visual displays within trials were identical. In auditory-detection tasks, each observation

contains different random noise components. In conceptual-discrimination tasks, observations consisted of random samples of material drawn from a specified source. In a sense, then, within-trial displays in the tasks where the model has proven accurate were nonredundant. In contrast, in the present task, within-trial observations were somewhat more redundant than in the auditory and conceptual tasks.

The two differences noted above, e.g., interference from the complex responses and "redundancy," provide a possible explanation for the present failure of the integration model. It was noted previously that, in the research cited here, the integration model proved effective only with complex response conditions. In the present study, the additional demands upon the visual system required by complex response conditions may have prevented these conditions from improving integration. Further, there is the possibility that motivation was involved. Specifically, the relatively redundant nature of the displays may have reduced the attention that Ss gave to additional displays.

To summarize the results of both experiments, integration fell far short of theory for observations beyond the second, and sometimes for the second observation as well. Requiring additional responses appeared to degrade discrimination. In both respects, the tilt-discrimination results differ from auditory-detection and

Table 4
Mean d' and d' Ratios for Multiple Response Groups

Number of Stimulus Observations	O_1	O_2		O_3			O_4			O_5		
	d'_1	d'_2	d'_2/d'_1	d'_3	d'_3/d'_1	d'_3/d'_2	d'_4	d'_4/d'_1	d'_4/d'_3	d'_5	d'_5/d'_1	d'_5/d'_4
Certainty												
1	1.00											
2	1.07	1.06	.99									
3	1.32	1.38	1.05	1.48	1.12	1.07						
4	1.60	1.64	1.03	1.59	.99	.97	1.72	1.08	1.08			
5	1.08	1.52	1.41	1.62	1.50	1.07	1.84	1.70	1.14	1.84	1.70	1.00
No Certainty												
1	1.40											
2	1.44	1.39	.99									
3	1.48	1.51	1.02	1.59	1.07	1.05						
4	1.89	1.93	1.02	1.93	1.02	1.00	1.89	1.00	.98			
5	1.24	1.63	1.31	1.67	1.35	1.02	2.01	1.62	1.20	2.04	1.65	1.01

conceptual-discrimination results. Possible post hoc explanations involve the lack of external variability in the tilt stimuli and the visual demands made by the act of responding. It is suggested that Ss are less responsive to additional stimulus presentations when external variability is lacking, and that the visual performance involved in making responses may hamper the visual performance involved in stimulus discrimination.

REFERENCES

ELLIOTT, P. B. Tables of d'. In J. A. Swets (Ed.), *Signal detection and recognition by human observers: Contemporary readings.* New York: Wiley, 1964.

GREEN, P. M., & SWETS, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

SWETS, J. A., & BIRDSALL, T. G. Deferred decision in human signal detection: A preliminary experiment. Perception & Psychophysics, 1967, 2, 15-24.

ULEHLA, Z. J., CANGES, L., & WACKWITZ, F. Integration of conceptual information. Psychonomic Science, 1967, 8, 223-224.

ULEHLA, Z. J., HALPERN, J., & CERF, A. Integration of information in a visual discrimination task. Perception & Psychophysics, 1968, 4, 1-4.