Otis Dudley Duncan
Hollingshead Lecture
Yale U. April 20, 1982

unpublished

Dudley Duncan
4523 Auhay Dr
Santa Barbara, CA 93110

RASCH MEASUREMENT AND SOCIOLOGICAL THEORY

Have you ever pondered the ambiguity of "and" in titles? Here I mean, "Rasch Measurement, <u>a Challenge to</u> Sociological Theory." The challenge is to take seriously a measurement model that is attractive in the light of commonly observed patterns in data and also for its fundamental logical and statistical properties. Taking it seriously will mean exploring carefully the conceptual consequences of the assumptions that <u>all</u> responses are probabilistic and that it is possible to separate the measurement of subjective personal traits (such as attitudes) and the measurement of social objects (such as questionnaire items or social entities or social values). The style of theorizing I want to see imitated is beautifully exemplified in the classic papers by P. E. Converse that introduced the concept of "non-attitudes" and pioneered in the mathematical modelling of responses to survey questions. After describing as nontechnically as possible some rather profoundly technical material, Converse wrote, "The substantive conclusion imposed by these technical maneuvers is simply that large portions of an electorate do not have meaningful beliefs, even on issues that have formed the basis for intense political controversy among elites for substantial periods of time." (p. 245, "The Nature of Belief Systems in Mass Publics.)

The technical maneuvers of the Rasch measurement model are somewhat different from those of the Converse two-sector model, but bear a close kinship to them, as I shall try to illustrate. The conclusion forced upon us if we accept the Rasch model is somewhat different from the one stated by Converse. We shall be accepting the premise that everyone who answers a question has an attitude, but it may well be true that "large portions of an electorate" do not hold very extreme attitudes on issues that exercise elites. Elites and the mass public may well have quite different <u>distributinns</u> of attitudes (even if their means coincide). But elite and mass alike will answer your questions as though using some kind of randomizing device. For the elites, perhaps, the devices will assume probabilities close to zero and one; for the masses, the probabilities will perhaps be toward the center of this range. That, at any rate, will be the explanation suggested by the Rasch model when we run into the kind of ostensible instability of response that was so interesting to Converse.

Let me first illustrate the kind of data that make <u>some</u> kind of probability
1A model for response a virtual necessity. In the top part of Figure 1 we see a small set of panel or re-interview data. The first interview was in the Spring or Summer; the second interview followed it by some 1 to 8 months. The data were collected in the Detroit Area Study. The question (YG) reads:

> Which statement do you agree more with?
> --The younger generation should be taught by their elders to do what is right.
> --The younger generation should be taught to think for themselves even though they may do something their elders disapprove of.

The salient feature of the table-- the substantial amount of apparent turnover -- is highlighted if we compare it with results from a different re-interview study
1B conducted by the Bureau of the Census in 1950. We have here (lower part of Figure 1) a classification of urban occupied dwelling units by tenure (renter-occupied versus owner-occupied) in the census and in the post-enumeration survey. There are a few errors (apparent changes in tenure, which in concept are ruled out by referring the classification to the census date); but there is essentially perfect agreement of the two classifications. The results for the YG classification could, of course, be due to the fact that some 25% of the sample actually "changed their minds." But it is not easy to say why that should be happening. We were interviewing the same people with the same minds and we know of no short-run shocks to the social system that would produce mental instability in a quarter of the population.

2 In a second example, Figure 2, we again see a large fraction of respondents (3/8) "changing their minds." But the table is actually restricted to the 85% of all respondents who said they still "felt the same" about garbage collection as they had at the time of the first interview.

Another kind of problem sometimes noted with survey data is that a non-negligible fraction of respondents seem to contradict themselves within a single interview. Ask two questions on the same issue and you will find a lot of people "changing their minds" in <u>both</u> directions. The tabulation in Figure
(3) 3 from the 1977 General Social Survey relates to these two questions:

> Q. 25: Do you approve or disapprove of a married woman earning money in business or industry if she has a [redacted] husband capable of supporting her?
>
> Q. 30: If there is a limited number of jobs, do you approve or disapprove of a married woman holding a job in business or industry when her husband is able to support her?

It is readily understandable that when it is "easier" to disapprove -- as in Q. 30, with its additional qualification concerning employment conditions -- a larger number of respondents will do so. But what is the explanation of the frequency of 58 for the No-Yes combination? We note that according to a well known procedure for scaling attitude items, the Guttman scale, this frequency is supposed to be zero. In practice, of course, one expects "error" in a Guttman scale, although the scale model as such says nothing about where the error comes from. The Rasch model is very much like Guttman scaling with an explicit mechanism producing such ostensible "errors," to wit, the assumption that response is probabilistic, not deterministic.

A third kind of problem that has captured much attention is the so-called attitude-behavior (A-B) problem: why do attitudes predict behavior imperfectly, sometimes
(4) very imperfectly indeed? In Figure 4 ~~with~~ we have three samples showing imperfect A-B correlations (they are actually rather stronger correlations than those one often sees in data studied to shed light on this issue). The similarity of the three samples with respect to the amount of inconsistency between intention to vote and actual vote is striking. Let me call attention, in particular, to the close similarity (identity within limits of sampling error) of the <u>ratios</u> of the two off-diagonal cell counts in the three tables.

Examples like these are certainly familiar, but it is worth asking why we do not continually remind ourselves of these awkward situations. The answer is that they do not arise in any obvious way in the usual cross-sectional analysis of a single response. It takes special strategies to bring them to light and, more importantly, to do something with them. I shall try to make the point with some simple simulations.

(5) We begin by contrasting two situations, Figures 5 and 6. In both situations the probability, p, of getting the shaded pattern on one trial is 7/12, or about 58%. In Figure 5 we assume each respondent carries just one pellet, and on any trial will tell you whether it is the shaded or clear pellet. In Figure 6 we
(6) assume each respondent carries a spinner (demonstrate) and on each trial will spin it with probability p of getting the shaded pattern. With simple random sampling the relevant statistical distribution in both situations is the binomial (nicely approximated by the normal for large samples). To tell these two situations apart we might ask some other question which could be correlated with type of pellet in the first situation only. But better yet, we take a second observation on the same respondents, that is, we do a re-interview. The expected outcomes
(7) (sampling error ignored) are in Figure 7. The top contingency table illustrates perfect association, the bottom one complete independence.

In actual re-interview studies we seldom see either extreme (the census study was approximately the top case). This leads Converse to suggest that real
(8) life is a mixture of two sectors as in Figure 8, where part of the population carries pellets and respond deterministically, the remainder use the spinner and answer "at random." With such a mixture there is a wide range of possible data patterns for the interview versus re-interview table; Case III in Figure 7 illustrates one possibility.

As Converse noted, while the amount of apparent "mind changing" in
2-wave panel data may be great enough to suggest the operation of a 2-sector
model, actually to test his model and to estimate its parameters requires three
waves, that is, two re-interviews. Before considering what might happen with
this rare kind of survey design, let us consider an alternative to the
2-sector model, a model with arbitrarily many sectors, with all respondents using spinners
but with different probabilities. The Rasch model actually treats probabilities
as varying continuously among respondents; so in principle there are infinitely
9 many different spinners, but for sake of illustration, in Figures 9 and 10 I
assume just 7 different probabilities. There is, therefore, a generally unknown
population distribution of p_i (probability of shaded for the i-th respondent).
The two figures illustrate somewhat contrasting distributions. In Figure 9, most
10 R's have p_i close to $\bar{p}$ = 7/12. In Figure 10, most R's have extreme p_i, far
above or below $\bar{p}$ = 7/12. Note the identity of $\bar{p}$ in the two simulations. For
a single response, the binomial distribution with $p = \bar{p}$ is still the appropriate
model under simple random sampling. But if we simulate the re-interview study in
11 these two populations, we get the rather different results noted in Fig 11 for Cases IV and
V. For comparison, Case III, simulating Converse's 2-sector model, is repeated.
Note that it would be easy to make the 2-sector and the many-sector simulations
come out just the same if we cared to.

In the Rasch model, the probabilities themselves are not the fundamental person parameters. This is perhaps fortunate, for it is difficult to imagine estimating directly each person's probability by taking a random sample of, say, 250 of his responses to the same question, asking him to "erase his mind" (as Lazarsfeld used to say) between each asking of the question. Instead, the model has these postulates:

(1) Response (including actions or reports of behavior, if we want to bring them within the scope of the model) is probabilistic.

(2) For any item, the response to it is supposed to depend (probabilistically) on an unobserved disposition or tendency of the respondent, and nothing else.

(3) If there are two items/responses to which depend on the same individual disposition, the probability of a positive response to each is governed by the item parameter as well as the person parameter, and by nothing else.

Thus, we take response to be a function of both the <u>person</u> and the <u>item</u> or object; and the definition of the relevant object may include aspects of the situation or context in which it occurs, so that voting in 1948 (in a presidential election) is a different item from voting in 1950 (in a congressional election). To separate these two factors, the item parameter and the person parameter, in a carefully stipulated sense, according to Rasch, dictates the mathematical form of the model. He was seeking "specific objectivity," which means that

> the comparison of any two subjects can be carried out in such a way that no other parameters are involved than those of the two subjects—neither the parameter of any other subject nor any of the stimulus parameters. Similarly, any two stimuli can be compared independently of all other parameters than those of the two stimuli.

What is sometimes called the "single parameter logistic" model, Rasch claims to have proved, is the only one that achieves specific objectivity in this sense.

The geometric representation of the model is particularly simple if we transform the response probability to y_{ij} = log odds on positive response to item j by person i. The model is then

$$y_{ij} = \alpha_j + \theta_i$$

12 Figure 12 illustrates for the case of just two items measuring the same disposition (or latent trait), i.e., the unobserved magnitude θ. The two lines are parallel and run at a 45° angle to the axes. Persons and items are measured on the same scale.

A simple algebraic derivation, which I shall not demonstrate here, turns up one important property of the model, the invariance of item parameters across groups or strata of a population -- <u>any</u> groups. A test of the model, therefore, is obtained by considering groups that observably have different response probabilities and therefore, presumably, differ in their distributions on the latent trait (or disposition). We must find a special kind of invariance across such groups if the model holds for two or more items. An example was given earlier, Figure 4. "Intention" and "Actual" vote are the two items. The electorate in three years comprises the set of three groups. The required invariance was noted when we looked at these data -- the ratios of counts in the two off-diagonal cells were the same (apart from sampling variation) [redacted] across the three groups. We note that as a matter of statistical method, this test is nothing more than the chi-square test of the hypothesis of independence in a certain 3 × 2 table (to be sure, a table constructed in a special way). [redacted] I have carried out such tests within each of these 3 years for a variety of test factors that are plausibly considered causes of the disposition to vote: age, sex, color, education, income, occupation, party identification, and others. In most tests the required invariance is sustained.

Notice that in this work I am not specifically concerned with the correlation (association) of intended with actual voting (the A-B correlation). That is NOT what I test or analyze. [redacted] Nevertheless, if the Rasch model applies, it explains why attitudes do not perfectly predict behavior. The reason is that both A and B are probabilistically related to the underlying dispositions. In his splendid essay on disposition concepts, MOrris Rosenberg does not say precisely this. What he does say is

> The "circumstances" that permit, allow, or stimulate the conversion of dispositions into action are ... extremely varied. They may be thought of as cues, stimuli, triggers, releasers, necessary or sufficient conditions, and so on. Lacking knowledge of the stimuli, opportunities, environmental conditions, resources, anticipated consequences, or role contexts, the ability to predict specific behavior from general dispositions is usually weak.

In less elaborate phrasing, "many factors" affect the connection between dispostion and response or action. [redacted] A synonym for "many factors" is "chance." The consequences of an explicit chance component in response have been demonstrated in our earlier simulation of the responses to two items (or the same item used twice). The model also explains why attitudes do not perfectly predict attitudes and behavior does not perfectly predict behavior.

The same kind of explanation is afforded by [redacted] (1) the Converse 2-sector model, as generalized by Leo Goodman in a (2) specific modification of the Guttman scale model, and also by (3) any of the Lazarsfeld latent-class or latent-structure models and by (4) a variety of non-Rasch latent-trait models. Differences among these models provide a fine workout for specialists in measurement and models; the issues as between the alternatives are very much open. I have made just a little progress on two topics: (1) In regard to Rasch versus Converse, I have found that it takes four items, or a 4-wave univariate panel study to tell the two models apart. For three items they [redacted] yield identical fits to the data. With 4 items the models differ by a single degree of freedom, which affords a nice test of the supposed superiority of Rasch (if it should fit the data) over Converse. (2) In regard to Rasch versus Lazarsfeld's latent dichotomy model, I have a couple of examples where both models fit, so that it is possible to regard Lazarsfeld as a Rasch model with a highly constrained distribution of θ_i or Rasch as a Lazarsfeld model with constrained response probabilities. But there should be many more examples where one of these models but not the other gives a satisfactory account of the multivariate data.

As I have indicated, if the Rasch model applies, it satisfactorily explains both the "non-attitude" phenomenon noted by Converse (apparently random changes in response) and the "Attitude-Behavior correlation" problem. A third [redacted] bonus of the model pertains to its separation of the disposition (postulated to be an attribute of the individual) from the object or occasion (which the sociologist will want to insist is likely to be a collective representation or a matter of consensual definition). It then becomes conceivable that we could operationally distinguish between:

(a) a change in the population distribution of dispositions $f(\theta_i)$; and

(b) a change in the social evaluations or meanings of various objects of those dispositions, that is, α_j.

The distinction between these was, of course, one of the main points of Thomas and Znaniecki's "Methodological Note." Although these authors later modified their ways of looking at human attitudes, Thomas, for one, continued to insist on the "separation of attitudes and values, or psychological sets and tendencies to act, on the one hand, and the external stimuli to action on the other." The distinction was paraphrased thus by Fleming, "A value was any phenomenon that elicited a human response, an attitude was ... [redacted], a favorable or unfavorable disposition 'toward something' in the realm of values." Thomas and Znaniecki were not, of course, quantitative sociologists and they were not thinking of measurement as the first order of business. But the overwhelming advantage of a focus on formalism is that it frees you from unwanted connotations of necessarily unsatisfactory nomenclature. Thus, in this lecture I have deliberately used a number of different terms as sensitizing concepts when referring to what it is that we attribute to the person and what it is in the outer world that releases his observable response or action. Mathematically, the former is the location of the person [redacted] on the X-axis of my diagram (Fig. 12), and the differences among objects, values, items, social entities or whatever are represented by different intercepts of the item curves on that axis.

To follow up on this suggestion and actually achieve the separation of the person and item parameters may require renewed attention to the panel method, about which some methodologists have recently expressed a certain disenchantment. In the classic 2W2V (or 16-fold) panel table, each of the two variables provides a sample stratification that may be used to test whether the two measurements on the other variable behave like two items in a Rasch scale. A review of published 16-fold tables has provided examples in which both variables, or one, or neither variable passes this test. Failure of the model presumably means either (a) some flaw in the study design (which could include "contamination" of response to one item by responses to the other); (b) a change in the "meaning" of one or more items for some portion

of the population; or (c) an appreciable shift in individual dispositions. If we rule out the first two alternatives with adequate arguments we are left to conclude that there was a shift in the dispositions, with at least some people reorganizing their attitudes. Upon detecting such a shift, we may be inclined to attribute it to some known change in the milieu causing such a restructuring of dispositions. An inference of this kind is just as hazardous with the Rasch model as without it. But results obtained in testing the model at least tell us in a reasonably definitive way whether we actually have something to be explained. An example where we do NOT have such a revision of individual attitudes to explain is taken from the
13 1948 Elmira study (Figure 13).

The authors (Berelson, Lazarsfeld, and McPhee, 1954, p. 264) did not, of course, miss anything so striking as the shift in salience of class issues: "The image of Truman did not change, but the image of what was important in the campaign ... did change to a dominance of socioeconomic issues." But the data do not support their inference as to a causal influence which they seem to conceive of as operating on the dispositions of individual Democratic voters toward the candidate: "... saliency of class issues makes more of a difference in the later image of Truman than the other way round." On the contrary, the model for quasi-independence in the table (4 corners excluded) fits very well: $L^2 = 3.55$, d.f. = 5. This model says that there was no inter-wave change in the dispositions indicated by either item, that the 2 items measure different (though possibly correlated) dispositions in Rasch's sense, but that items are free to change parameter values between waves. The change in item parameter was significant for salience of class images, not so for image of Truman. On my interpretation, the 1948 campaign (or other developments during the June-August period) shifted the social definitions of the issues at stake (that, among other things, is what a campaign is supposed to do) but did not shift the individual dispositions reflected in either of the two items used in this analysis.

The data from Voting pertain only to Democratic voters and we are not shown a similar table for Republicans. But I will refer to the hypothetical possibility that one or more of the findings just noted does not apply to them. Suppose, in particular, that for Republicans class issues were salient all along and there was no change for them. In technical terms, the item parameters were not the same for Democrats and Republicans. In effect, the two parties comprise disjunct universes of discourse for the items used in this study. Here is a fourth bonus of the Raach model. It gives us a definite criterion for deciding when "questions mean different things to different people," if "different people" means "different groups of people" and we are willing to commit ourselves as to what the relevant groups are. I am conjecturing, therefore, that we have as much to learn from ostensible failures of the Rasch model as from its successes, if any, in our domain, valuable as successful attempts to create Rasch scales surely would be. Wasn't it true that geneticists learned a great deal from failures of the elementary Mendelian models, after the discipline had become convinced of the truth of Mendel's basic postulates? Only if we take Rasch seriously can we gain something from the demonstration that the measurement model does not hold in particular instances.

[redacted]

YOUNGER GENERATION

FIRST INTERVIEW	RE-INTERVIEW DO RIGHT	RE-INTERVIEW THINK
DO RIGHT	58	27
THINK	19	78

N = 182

Fig

URBAN OCCUPIED DWELLING UNITS

CENSUS	RE-INTERVIEW RENT	RE-INTERVIEW OWN
RENT	87	3
OWN	2	90

N = 182 (SIMULATED)

Fig

SATISFIED WITH GARBAGE COLLECTION?

FIRST INTERVIEW	RE-INTERVIEW WELL	M/L	NOT
WELL SATIS.	84	25	3
MORE OR LESS	15	12	4
NOT AT ALL	8	6	7

TAB FOR 164 R's (85% OF TOTAL) WHO "FEEL THE SAME NOW AS ____ MONTHS AGO"

L^2 for Rasch model = 5.29

d.f. = 3

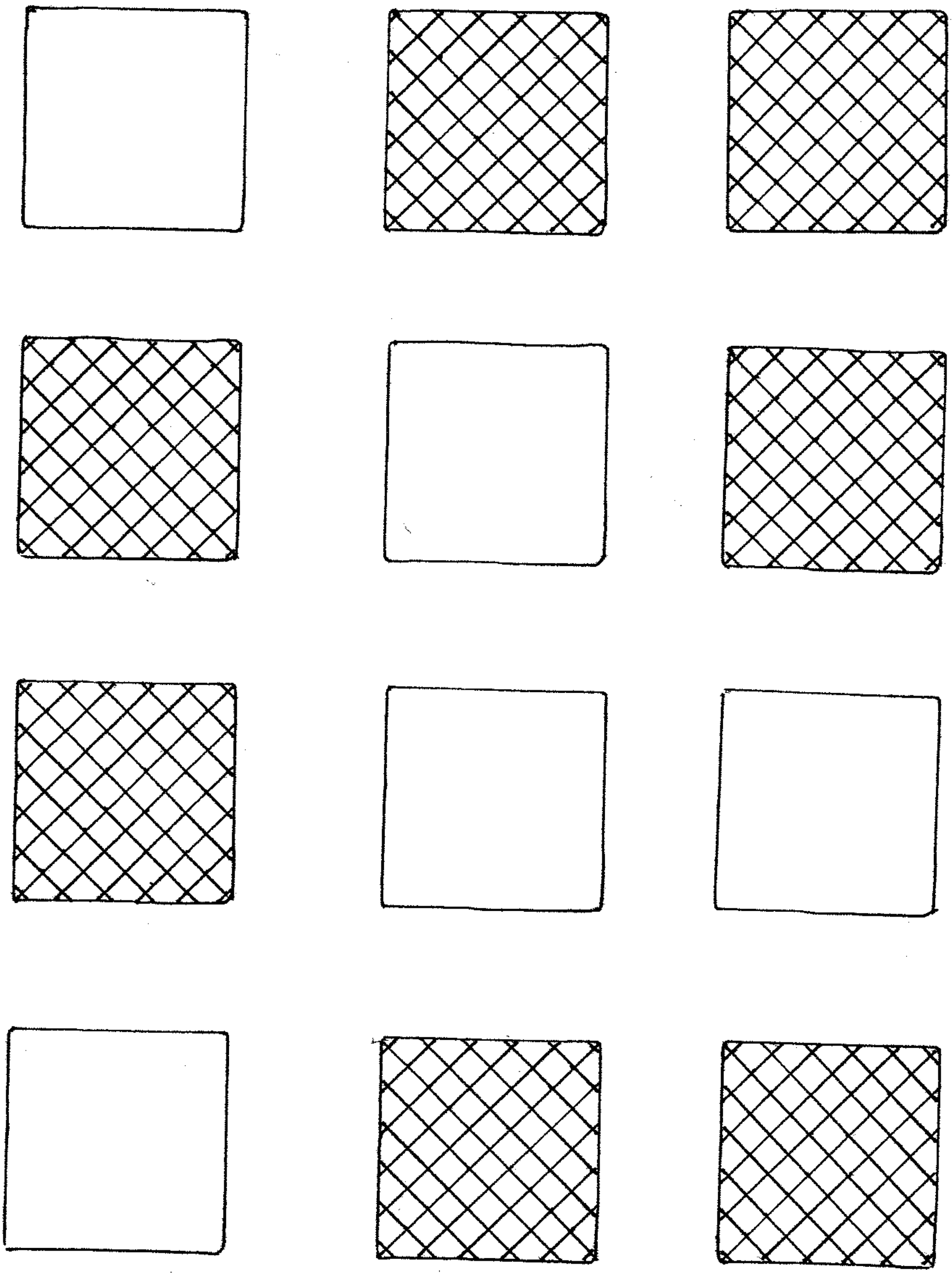
APPROVE WOMAN WORKING?

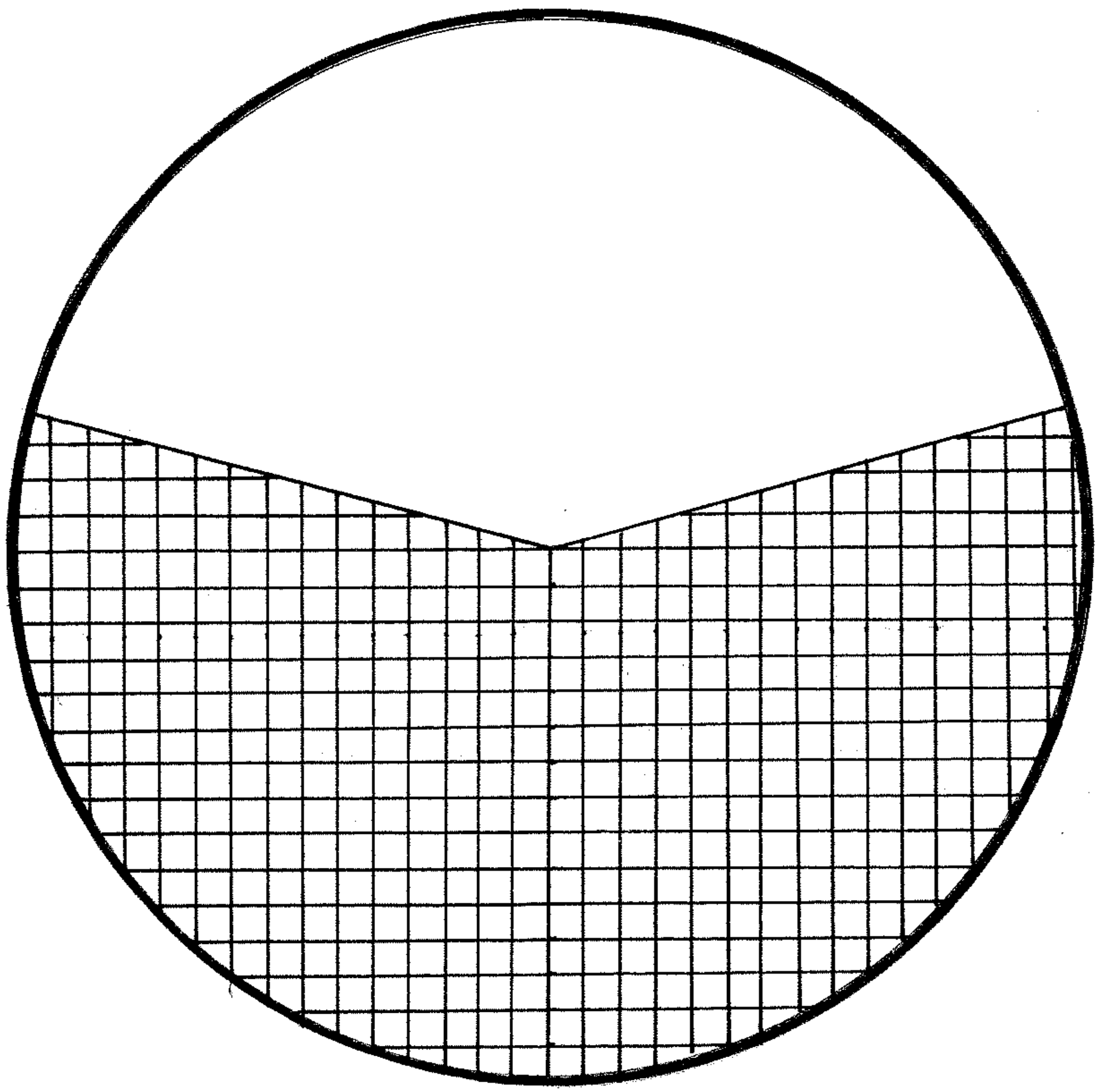
Q. 25	Q. 30 YES	Q. 30 NO
YES	478	502
NO	58	437

Q. 25: 66 %

Q. 30: 36 %

INTENTION	ACTUAL VOTE	ACTUAL NOT VOTE
1976		
VOTE	1348	175
NOT VOTE, DK	44	342
1972		
VOTE	1605	248
NOT VOTE, DK	57	373
1950		
VOTE	899	191
NOT VOTE, DK	56	202



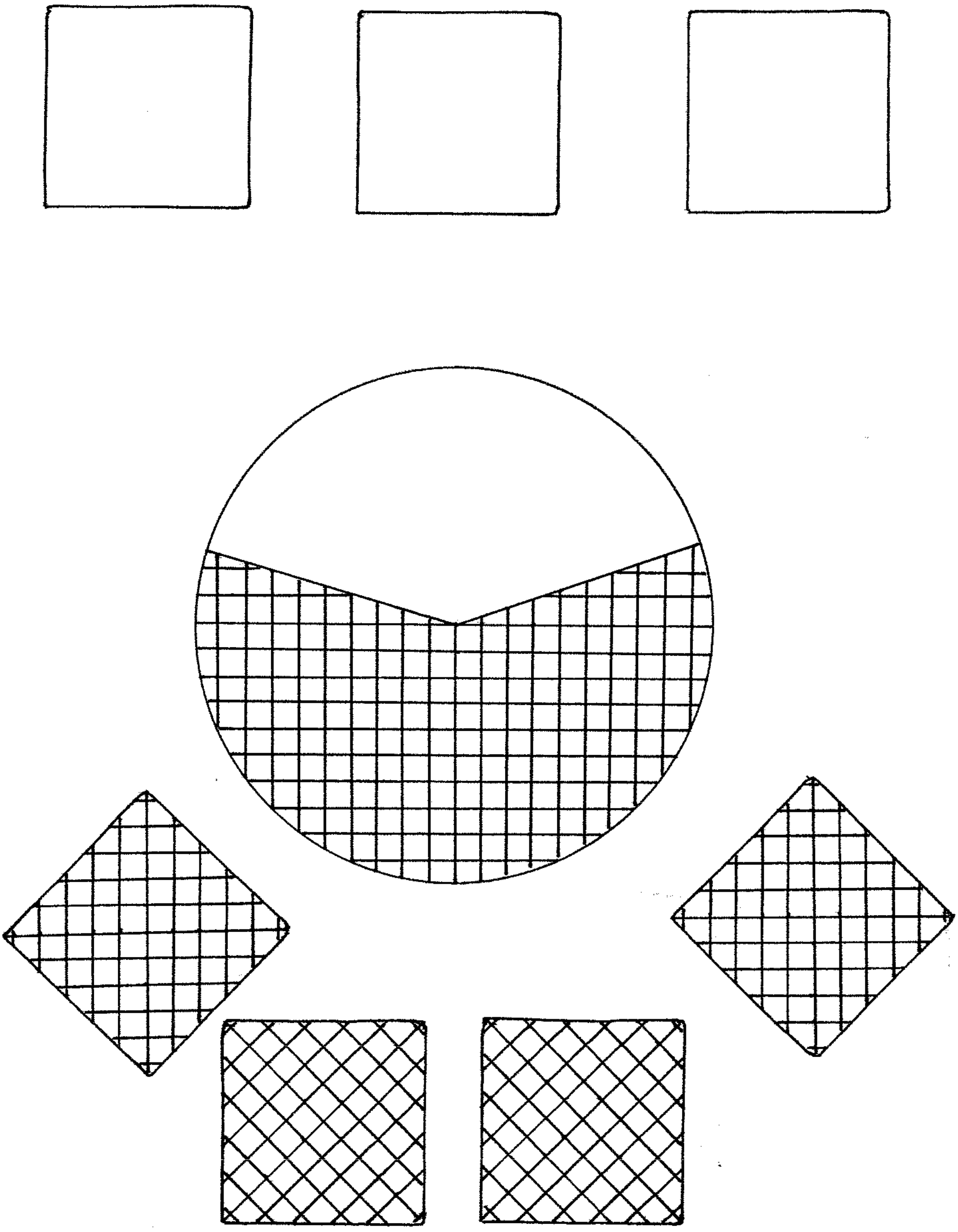


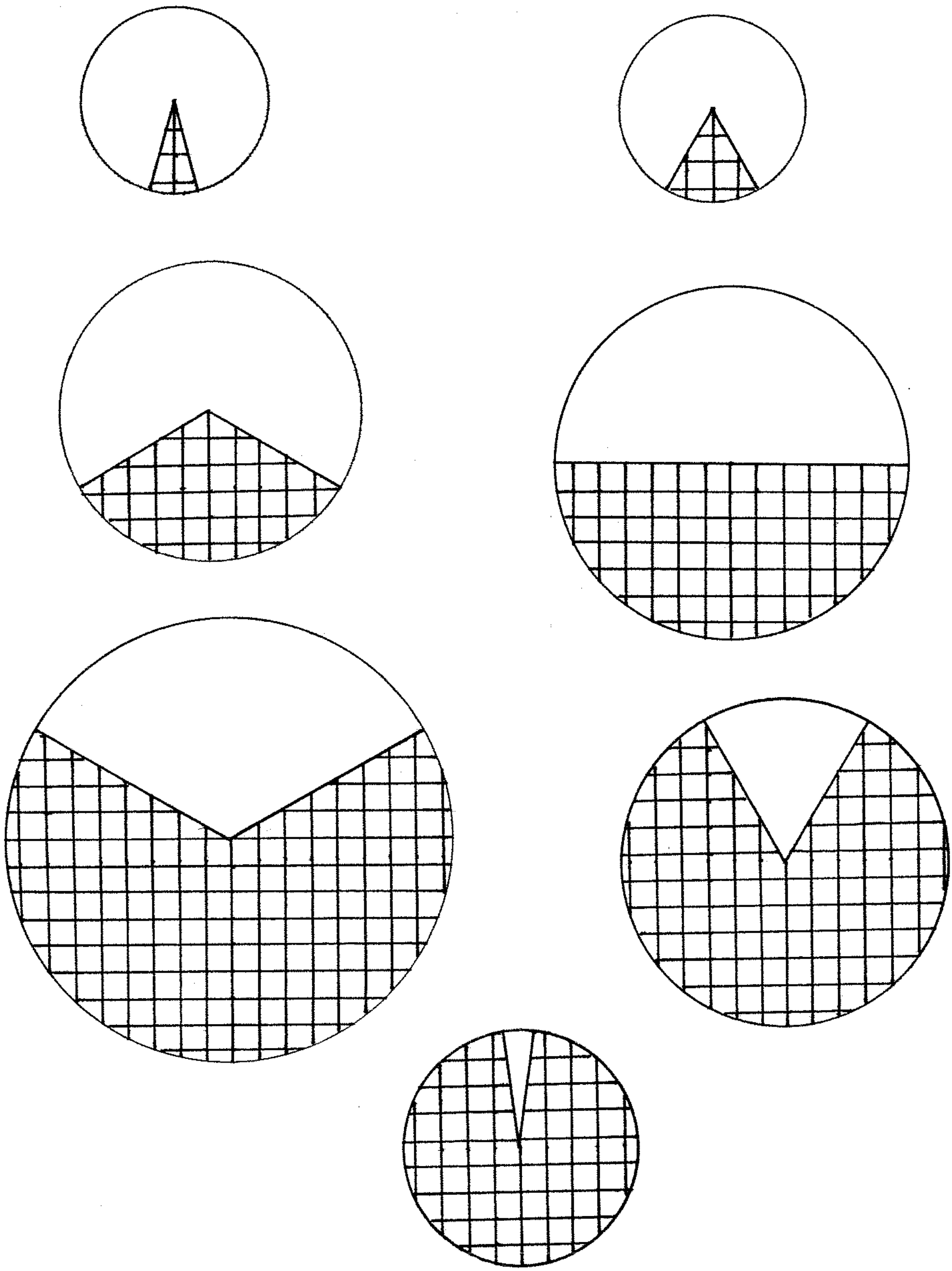
(N = 144)

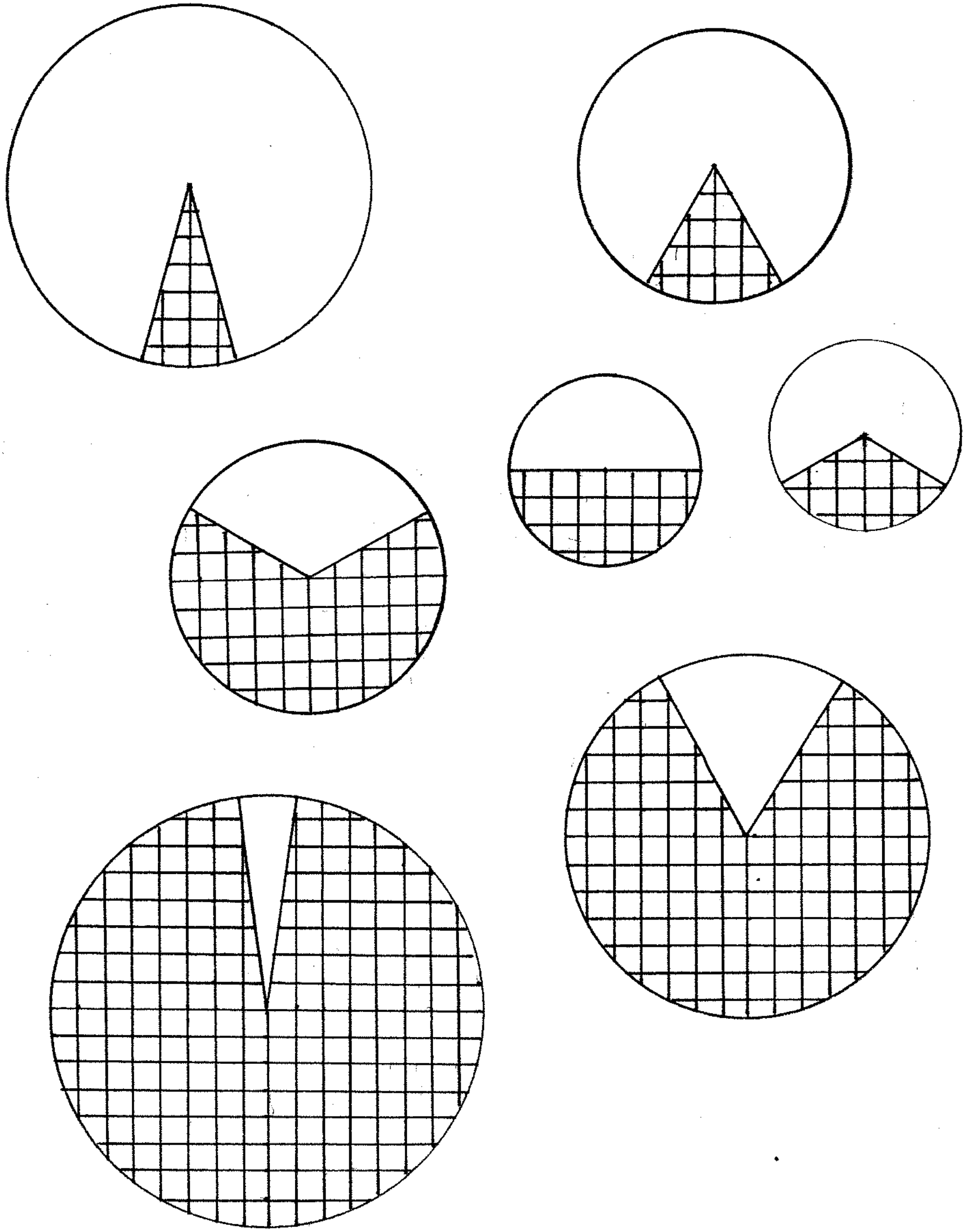
		□	▦
I.	□	60	0
$p = 7/12$	▦	0	84

		□	▦
II.	□	25	35
$p = 7/12$	▦	35	49

		□	▦
III.	□	46	14
$p = 7/12$	▦	14	70



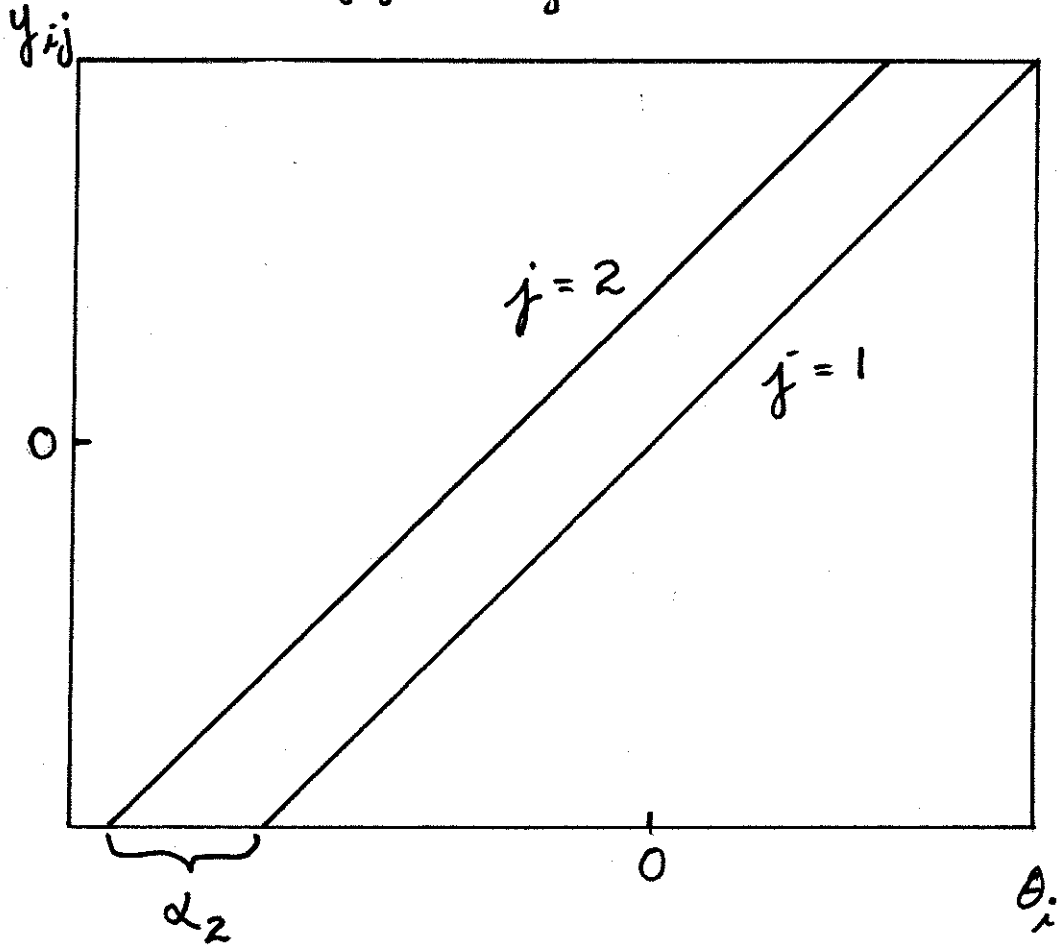




(N = 144)

		□	▩
III.	□	46	14
$p = 7/12$	▩	14	70
IV.	□	34	27
$p = 7/12$	▩	27	56
V.	□	43	17
$p = 7/12$	▩	17	67

$$y_{ij} = \alpha_j + \theta_i$$



y_{ij} = log odds on positive response by respondent i to item j.

θ_i = person parameter

α_j = item parameter ($\alpha_1 = 0$)

CLASS ISSUES SALIENT?	IMAGE OF TRUMAN				
JUNE → / ↓ AUG	FAV FAV	FAV UNFAV	UNFAV FAV	UNFAV UNFAV	TOTAL
Y Y	20	2	6	7	35
Y N	8	1	3	6	18
N Y	52	14	16	37	119
N N	54	23	19	60	156
TOTAL	134	40	44	110	328

$$L^2(M_5) = 3.55, \quad d.f. = 5$$

$$\log(\hat{b}/a) = 1.889 \ (.253)$$

$$\log(\hat{d}/c) = .095 \ (.218)$$