THE PROCESS OF RESPONDING TO PERSONALITY ITEMS: INCONSISTENT RESPONSES TO REPEATED PRESENTATION OF IDENTICAL ITEMS

JAMES A. BOND

Department of Psychiatry, University of Michigan, 900 Wall Street, Ann Arbor,
MI 48105, U.S.A.

Summary—A promising approach to understanding the processes involved when subjects respond to personality items is provided by the investigation of the causes of *inconsistent* responses when subjects answer the same item on two occasions. Among these causes are the properties of the item. Previous item research focused almost exclusively on properties which are not highly specific to the item, such as endorsement rate (ER) and social desirability scale value (SDSV). Although past studies found that items with 'extreme' SDSVs and/or ERs elicit fewer inconsistencies, these studies ignored more item-specific properties such as item content and item ambiguity. The present study demonstrates that contrary results regarding consistency may be obtained when more item-specific properties are taken into consideration. These results are interpreted as evidence that certain kinds of item content can increase the indecision and conflict that characterize some subjects' response processes.

With so much discussion of social desirability and acquiescence in the 1960s one might expect that substantial progress had been made in understanding the determinants of an individual's response to a personality inventory item. However, as Dahlstrom, Welsh and Dahlstrom (1975, p. 128) noted in their monumental review of MMPI research, "our understanding of the actual processes mediating the behavior of a test subject is disappointingly meager". Nonetheless, since the 1960s, some promising steps have been taken toward illuminating the psychological processes involved when an individual responds to a self-descriptive item (Angleitner, John and Loehr, 1985; Bond, 1984; Cliff, 1977; Fekken and Jackson, 1983; Kuncel, 1981; Markus, 1977; Rogers, 1973, 1974).

Perhaps the most important conclusion to be drawn from recent research is that the response process is considerably more complicated than earlier discussions suggested. There are a wide variety of influences that are potentially involved, though no single causal influence need be at work in any particular case. Subjects may allow their responses to be influenced by a conscious wish to appear in a socially desirable light, or on the other hand, by a conscious wish to appear more disturbed than they really are. Some respondents are too careless to even read the content of the items prior to answering. Responses are sometimes influenced by complicated unconscious wishes, e.g. a tendency to exaggerate the intensity of distressing experiences may lead to an elevated profile on a test like the MMPI, and yet this is not the same process as the more frequently discussed conscious wish to 'fake bad'. Some responses may be influenced by the mood or 'state' of the respondent at the specific time that answers are given (Coppen and Metcalfe, 1965; Kendell and Di Scipio, 1968). It is even possible that some respondents read the items carefully and give candid responses based on their own self-knowledge!

Note that this list of potential influences only involves individual differences that may characterize the *respondents* (and is not even exhaustive of these). Properties of the *items* to which responses are given, can also influence the response process. Some of these personality item properties are: the ambiguity of item phrasing, the particular type of content with which the item deals, the endorsement rate of the item, the social desirability scale value of the item, and the position of the item in the order in which all the items are presented. Angleitner *et al.* (1985) have persuasively emphasized the necessity for careful investigation of the syntactic and semantic properties of personality items because of the effects these properties may have on the reliability and validity of subjects' item responses.

In addition to variation in respondent characteristics and variation in item properties, there are also potential *interactions* that make take place between specific classes of respondents and specific
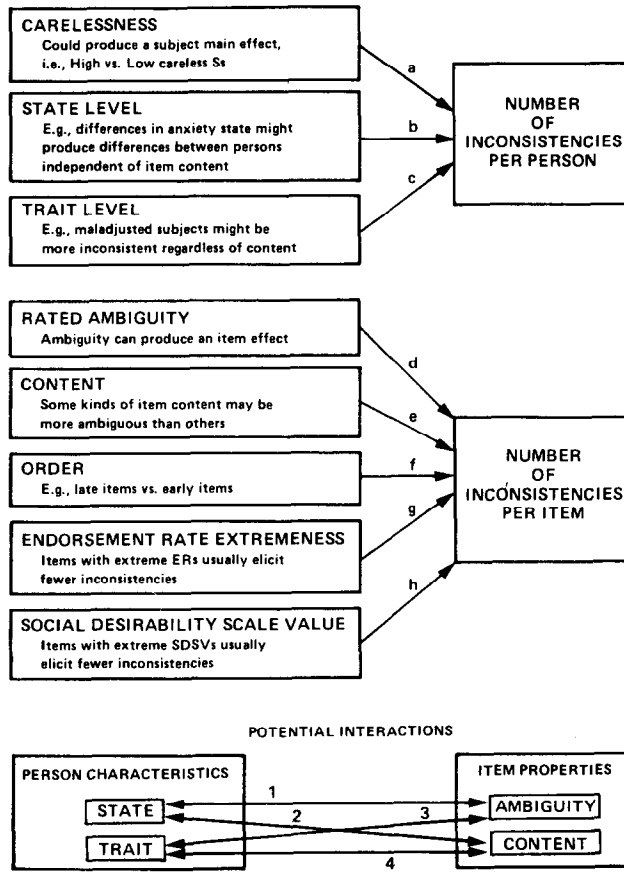
Fig. 1. Some attributes of persons and of items which may, alone or in interaction, influence the number
of inconsistencies elicited.

types of items. For example, if a respondent were in a depressed state, his/her response process
would more likely be affected by this state when the item content reflects mood-related
properties. Response to an item such as "I find it hard to keep my mind on a task or job" might
be more influenced by depressed mood than response to an item such as "I like mechanics
magazines".

One promising method of investigating the causal influences which may affect the process of
responding to personality items involves having subjects respond to identical items on two (or
more) occasions. Such data allow investigation of the causes of 'inconsistent' responses. When the
same subject responds to the same item on two occasions, an inconsistency is a changed response
to that item. An understanding of why individuals change their responses to the identical item
stimulus from occasion to occasion can help us understand how they make up their minds on any
single occasion. Figure 1 is a graphic presentation of some of the possible influences on the
responding process as they might affect inconsistency of responses. We should like to know which
item properties elicit response stability and which elicit more frequent inconsistent responses. The
issues addressed in the present paper mainly concern the portion of Fig. 1 involving number of
inconsistencies per item.

It is usually *assumed* that personality items which elicit more stable responses are preferable
(Fekken and Jackson, 1983; Angleitner *et al.*, 1985). This may or may not be the case, depending
on the clarification of a number of complicated psychometric relationships. For example,
clarification is needed of the relationship between item stability and scale stability. If less stable
items do not necessarily produce substantially less stable scale scores then the desirability of
maximally stable items is called into question. Another relationship in need of clarification is that
between scale stability and scale validity. For example, Dahlstrom *et al.* (1975, pp. 175ff) argued
that low stability of test–retest scores for certain MMPI scales may reflect valid changes in

respondents which occured in the test–retest interval. Finally, there is the relationship between item stability and scale validity. Recall the debate over whether item subtlety and/or ambiguity, properties which can produce item instability, are or are not associated with greater scale validity (Meehl, 1945; Jackson, 1971). Although there are established views regarding these issues as they apply to abilities items, these views are often still controversial in their application to *personality* items. One implication of these controversies is the still open possibility that personality items which elicit only moderate stability may yet produce more valid scales.

What personality item properties might affect the number of unstable test–retest responses elicited? Previous investigators (Edwards and Walsh, 1963; Fekken and Jackson, 1983; Goldberg, 1963; Payne, 1974) found that personality items with extreme social desirability scale values (SDSV) and/or extreme endorsement rates (ER) elicited fewer inconsistent (i.e. unstable or changed) test–retest responses. 'Extremeness' of SDSV refers to the absolute distance of an item's SDSV from 5.00, the midpoint on the usual scale for SDSVs. 'Extremeness' of ER refers to the absolute distance of an item's ER from 0.50, the midpoint of the range for ERs (Payne, 1974). However, these earlier investigations did not consider any of the other properties of personality items which might well be expected to affect the number of inconsistent responses elicited, e.g. specific item content and item ambiguity.

The distinction between *specific* item content and the relative nonspecificity of item indices like ER and SDSV must be made clear. First, why is extremeness of ER a relatively nonspecific item property? Because to know that a set of personality items is answered 'True' by, say, 25% or less of a sample of respondents (i.e. has an 'extreme' ER), is to know very little about the personality items in question. There are a wide variety of reasons why subjects decline to endorse items, and these reasons concern the specific content of the items. Perhaps some of the items involve the self-ascription of psychotic experiences, some may involve the self-ascription of bad public table manners, some may involve the self-ascription of unrealistic virtuousness, while still others might involve the self-ascription of more common yet undesirable traits of personal maladjustment. In other words, a group of items that share the property of having extreme ERs may reflect a wide variety of specific item contents. An examination of the heterogeneous content of MMPI F scale items (Dahlstrom, Welsh and Dahlstrom, 1972, pp. 112ff) demonstrates this; these items were selected solely because their response distributions were extreme. Thus, the ER of an item tells us nothing specific about an item. ER is a nonspecific, aggregate (i.e. a proportion of a sample) item property.

Second, although SDSV does involve item content—because judges rate the social desirability of an item on the basis of its content—SDSV does *not* reflect any specific item content. This can be demonstrated by examining the heterogeneous content of the first seven MMPI items with SDSVs ≥7.00, i.e extremely desirable if answered 'True' (see Table 1). These items reflect an evaluation of one's father, one's socializing ability, one's freedom from sexual problems, one's churchgoing behavior, etc. They have little, if any, specific item content in common. Yet they all share a nonspecific item property, they have extreme SDSVs.

The present paper reports the results of an investigation of the effects of specific item content and item ambiguity, as well as SDSV and ER, on the number of inconsistencies elicited by selected MMPI items.

METHOD

As part of a larger investigation of the sources of inconsistent responding to personality items (Bond, 1984, 1986), two sets of MMPI items were selected: one whose content reflected

Table 1. First seven MMPI items with extremely positive SDSVs (when answered true)

Item No.	Item content	SDSV
3	I wake up fresh and rested most mornings	7.75
8	My daily life is full of things that keep me interested	8.28
17	My father was a good man	7.72
37	I have never been in trouble because of my sex behavior	7.49
57	I am a good mixer	7.15
95	I go to church almost every week	7.15
96	I have very few quarrels with members of my family	7.30

Table 2. First 10 MMPI items judged neutral with regard to maladjustment

Item No.	Item content
1	I like mechanics magazines
4	I think I would like the work of a librarian
12	I enjoy detective or mystery stories
25	I would like to be a singer
60	I do not read every editorial in the newspaper everyday
70	I used to like drop-the-handkerchief
77	I enjoy reading love stories
78	I like poetry
81	I think I would like the kind of work a forest ranger does
83	Any man who is able and willing to work hard has a good chance of succeeding

maladjustment and another whose content was relatively neutral with regard to maladjustment. The 83 maladjustment (MAL) items consisted of the social maladjustment (SOC), depression (DEP), and poor morale (MOR) content scales developed by Wiggins (1966). The 83 neutral (NEUT) items were selected by two independent ratings of the content of each MMPI item as to whether it was Neutral with regard to maladjustment, Possibly Neutral, or Non-Neutral. Since the MMPI includes a substantial minority of items reflecting preferences, beliefs, opinions, and relatively innocuous self-descriptions, it was possible to select 83 NEUT items. A list of the MMPI Group Form item numbers for these two sets of items is presented in the Appendix. The first 10 of the NEUT items are shown in Table 2 to provide the reader with the flavor of these items.

Subjects. The data reanalyzed here were collected by Goldberg (1978). For a fuller description of his procedure see his paper. Subjects were 94 male and 108 female University of Oregon undergraduates enrolled in an introductory psychology course. They responded to the MMPI in class on two occasions. Administrations were separated by a 4-week interval during which they responded to two other inventories. Instructions emphasized that the students should respond to each item 'as you feel today, regardless of how you may have answered in the past'.

RESULTS

Given past results indicating that personality items with extreme SDSVs and ERs elicit more inconsistencies, while items with moderate SDSVs and ERs elicit more inconsistencies, the first task was to examine the distributions of SDSVs and ERs for MAL and NEUT items (see Fig. 2). Examination of the distributions in Fig. 2 shows that NEUT items have more moderate and fewer extreme cases among both SDSVs and ERs than do MAL items. Table 3 presents the results of dichotomizing all four distributions into Extreme vs Moderate items and applying chi-square to determine whether the differences are statistically significant. NEUT items had more Moderate and fewer Extreme cases of both SDSV and ER than MAL items and these differences were significant at the 0.05 level.

These results, coupled with previous findings (Edwards and Walsh, 1963; Fekken and Jackson, 1983; Goldberg, 1963; Payne, 1974), might lead to the expectation that the MAL items should elicit fewer inconsistencies and the NEUT items more. However, it is of interest that just the reverse is true. A repeated measures ANOVA (Winer, 1971, pp. 261–273) showed that the difference between the number of inconsistencies per item on MAL items (Mean = 29.95, SD = 14.26) and the number of inconsistencies per item on NEUT items (Mean = 21.1, SD = 11.73) is statistically significant (obtained $F = 95.24$, while F at 0.01 (1,201) = 6.76). For some reason MAL items elicit *more* inconsistencies than NEUT items among this sample of normal college undergraduates.

Some further comparisons were done to illuminate this somewhat anomalous finding (see Table 4). Perhaps most important among these further comparisons was the selection of two item subsets consisting of 40 MAL items and 40 NEUT items, where each item was matched for ER with an item in the other subset. In other words, these two item subsets, though differing in item content, consisted of items with virtually identical ERs. Goldberg (1963) has emphasized the importance of an item's ER in contributing to the number of inconsistencies it might elicit, and matching items on ER represents a control over this potential source of variance. Given the distributions of SDSVs among these two 40 item subsets it was not possible to match them for SDSV simultaneously; however, the type of difference that exists (i.e. MAL items had more extreme
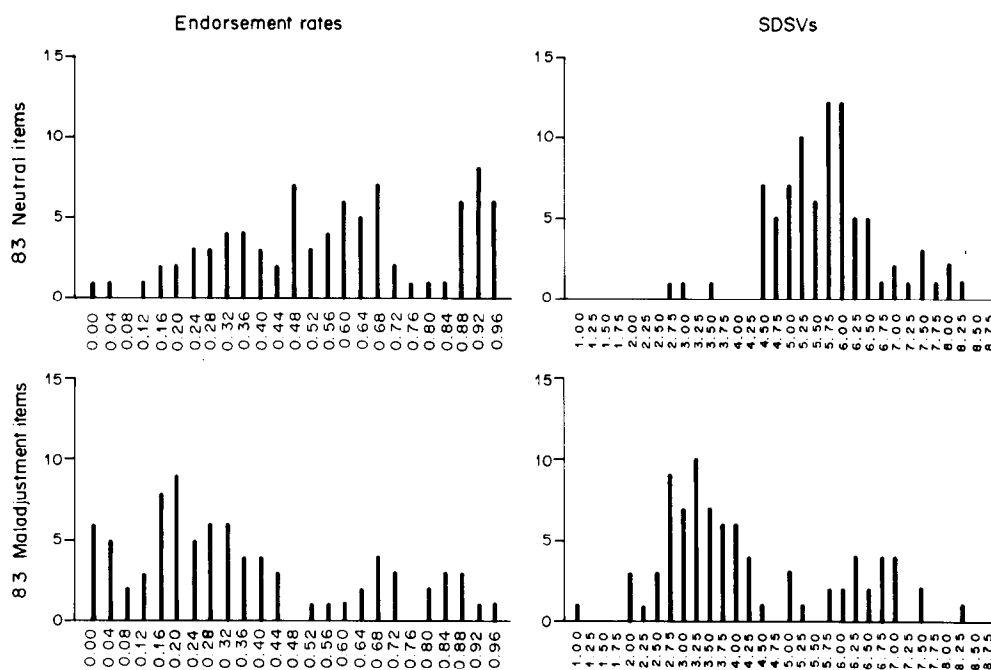
Fig. 2. The distributions of SDSVs and Endorsement rates for 83 Neutral and 83 Maladjustment items show that the Neutral items have more midrange cases on both parameters.

SDSVs, see Fig. 2), would be expected to produce more inconsistencies among the NEUT items as compared with the MAL items. However, a repeated measures ANOVA again produced a highly significant F ratio ($F = 85.25$, see Table 4 for Means and SDs). Although these two subsets of items are equivalent in several ways—they consist of an equal number of items (40), they have virtually identical ERs, they were responded to by the same 202 subjects on the same occasions, and they are parts of the same test (MMPI)—the 40 MAL items elicit significantly more inconsistencies on average than do the 40 NEUT items, among normal college students. And this though the differences in SDSVs between the groups should, on the basis of past research, have pulled in the direction of more inconsistencies among NEUT items.

Since the MAL items ask for more sensitive personal revelation (e.g. about anxieties, interpersonal problems, and personal morale), one might think that maladjustment items *per se* would elicit more inconsistencies than items whose content is relatively neutral with regard to psychopathology. The mean number of inconsistencies per item on Wiggins' 48 Psychoticism items (18.24), as shown in Table 4, indicates that this is not the case. Items reflecting psychotic level maladjustment elicit fewer inconsistencies on average (among normal college students) than any other item group shown.

Table 3. Dichotomized frequency distributions of SDSVs and ERs
for both neutral and maladjustment items

	Social desirability scale values*	
	Moderate (7.0 > SDSV > 3.0)	Extreme (3.0 ≥ SDSV ≥ 7.0)
Neutral items	73	10
Maladjustment items	59	24

	Endorsement rates†	
	Moderate (0.76 > ER > 0.24)	Extreme (0.24 ≥ ER ≥ 0.76)
Neutral items	50	33
Maladjustment items	35	48

*Chi-square (with Yates' correction) for SDSV frequencies = 6.25, 1 df, $P < 0.05$.
†Chi-square (with Yates' correction) for ER frequencies = 4.73, 1 df, $P < 0.05$.

Table 4. Means and SDs of number of inconsistencies per item among various groupings of
MMPI items

Item group	Mean no. of inconsistencies per item	SD
83 Nonpsychotic maladjustment items	29.95	14.26
83 Neutral items	21.10	11.73
48 Psychoticism items	18.24	14.20
All 566 MMPI items	26.25	14.75
40 Maladjustment items (matched for endorsement rate)	32.10	14.43
40 Neutral items (matched for endorsement rate)	21.50	11.64

Number of inconsistencies per item is based on a constant number of subjects, $N = 202$.

The comparison between 40 (nonpsychotic) MAL items and 40 NEUT items with matched ERs may be further illuminated by examining the 'ambiguity' ratings of these same items. Harris and Baxter (1965) collected ambiguity ratings on all MMPI items and these can be found in Appendix F of Dahlstrom et al. (1975). In their study 'ambiguity' meant the difficulty "you find in responding to the item for one or all of several reasons—the specific wording of the statement; the vagueness of the statement; the many alternative meanings; or, finally, the generality of the statement with regard to time place or circumstance". Having found that the 40 MAL items elicit more inconsistencies than NEUT items it would be of interest to discover whether these MAL items were perceived by an independent sample of normal college students as more ambiguous as well. The mean ambiguity rating of the 40 MAL items was 42.09 (SD = 10.27), while the mean for the 40 NEUT items was 28.48 (SD = 9.87). The difference is significant at the 0.01 level ($t = 2.68$, 78 df, two-tailed).

DISCUSSION

The present research found that significantly more MAL items had extreme ERs and SDSVs when compared with the NEUT items. Thus, past research, as cited above, would lead to the conclusion that MAL items should elicit *fewer* inconsistencies than NEUT items. However, the opposite result was obtained, MAL items elicited significantly more inconsistencies than did NEUT items. Further, a set of 40 MAL and 40 NEUT items with matched ERs showed the same anomalous result, MAL items with virtually identical ERs elicited more inconsistencies. This result was not due solely to the fact that MAL items requested potentially sensitive personal revelations while NEUT items did not, because Psychoticism (PSY) items, which also request potentially sensitive personal revelations, elicited the fewest inconsistent responses of any item subset investigated (see Table 4).

Why are the results of the present study contrary to those of past research? Recall our earlier discussion of the relatively nonspecific character of item properties like ER and SDSV, as compared with an item's specific content. Items that are similar because they have 'extreme' ERs (or SDSVs) may nonetheless have very different contents (see Table 1). Past research on personality items has been too dominated by attention to nonspecific properties like SDSV and ER, and has ignored more specific properties like item content. The approach taken in Payne's (1974) study illustrates this point. Payne divided all MMPI items into extreme and moderate categories based on ER and SDSV, *without regard to each item's specific content*. Such a division would place most psychoticism items in the extreme group and Table 4 shows that these items elicit few inconsistencies. The result obtained when testing for an association between extremeness (of ER or SDSV) and inconsistency will depend, at least in part, on the specific contents in the item pool being investigated. (It will also depend in part on the sample of respondents.)

Moreover, it is worth considering what sort of question is addressed by a study that demonstrates a correlation between extremeness and inconsistency without regard to the specific contents in the item pool used. Although it has often seemed that such studies were illuminating the response processes involved when individuals answer personality items, this is not the case. As Block (1965) and Norman (1967) pointed out, correlations between averages, like Edwards' (1953) finding that the average social desirability rating of an item is highly correlated with the proportion of subjects (another average) endorsing it, do not address a problem of great scientific interest, i.e. "What are

the determinants of an *individual's* response to a personality item, and how do these potential determinants combine to produce the answers given by certain classes of respondent to certain classes of item?" The typical respondent does not merely estimate the 'popularity' or social desirability of a 'True' response and answer accordingly. The typical respondent's answer is most likely a complex cognitive–affective–motivational response to the specific content of the item.

In the present study the primary criterion used in selecting the MAL, NEUT, and PSY items was their specific content. Why might the differences in content between MAL and NEUT items override the tendency for Extreme ERs and SDSVs to be associated with fewer inconsistencies among MMPI items? Perhaps one of the primary causes of inconsistency is the degree of *indecision* that the content of an item elicits in a respondent (Bond, 1986). (Note that this hypothesis directly addresses the response processes of the individual.) Three possible reasons that an item might elicit indecision are: (1) its degree of ambiguity as seen by the respondent, (2) its personal applicability or relevance as construed by the respondent, and (3) whether the answer that seems most accurately self-descriptive also arouses conflict in the respondent about openly admitting an undesirable trait.

Although the senses in which personality items may be 'ambiguous' are still not well understood (Angleitner *et al.*, 1985), their degree of ambiguity has long been believed important. The key words in personality items might be viewed in terms of a hierarchically organized semantic category model (Hampson, 1982), i.e. as representing behavioral categories which are ordered in terms of their level of abstraction from superordinate to subordinate. MAL items tend to involve relatively abstract traits like 'happiness' (e.g. "I wish I could be as happy as others seem to be"), while NEUT items more frequently involve concrete judgements (e.g. "I enjoy detective or mystery stories").

It was already noted that the content of MAL items was rated as significantly more ambiguous than the content of NEUT items by an independent sample of college students (Harris and Baxter, 1965). Presumably respondents experience more *indecision*, and are therefore more likely to give an inconsistent response on another occasion (Fekken and Jackson, 1983), when required to answer items involving more abstract and ambiguous terms—but, and this is a crucial qualification, only when respondents construe such items as coming 'close' to applying to themselves (Bond, 1984, 1986). For example, even though the PSY items were also rated as relatively ambiguous (Mean = 38.96, SD = 12.41, not significantly different from 40 matched MAL items but significantly different from 40 matched NEUT items, $P < 0.002$), they do not elicit many inconsistencies from normal college students because psychotic content is too extreme and irrelevant to such respondents and thus produces little indecision. However, the MAL items represent dimensions (e.g. social maladjustment, depression, and poor morale) which are more likely to overlap with normal personality traits like extraversion and optimism. Thus, they more frequently come close to applying to at least some normals and, in addition, the MAL items are relatively ambiguous. Finally, the MAL items are more likely to elicit some conflict over openly admitting an undesirable trait than are the NEUT items. Thus, the MAL items elicit more inconsistencies than NEUT items—even though the relative extremeness of the MAL items' ERs and SDSVs might suggest fewer inconsistencies—because the specific content of the MAL items (1) is more ambiguous, (2) more often comes close to applying to normal college respondents than, say, psychotic content, and (3) is more likely to elicit some internal conflict in a respondent.

Our hypothesis links inconsistency with indecision, and in turn links indecision with perceived item ambiguity, perceived personal applicability, and conflict about self-disclosure. One of the auxiliary findings of the Harris and Baxter (1965) study at least supports a link between perceived ambiguity and perceived personal applicability. They found that individuals who obtained higher pathological scores on some MMPI scales, i.e. found the items more personally applicable, also tended to rate the items in those scales as more highly ambiguous. Apparently when an item's content reflects sensitive personal revelation and is seen by a respondent as self-relevant, the item is also more likely to be seen as ambiguous.

It is worth calling attention to the fact that the hypothesized relationships between specific item content and specific type of respondent are interactional hypotheses (see 'Potential Interactions' in Fig. 1). Interactions of this sort most directly address the issue of personality item response processes.

The explanations offered in this study concern the *specific content* of a personality item: its degree of ambiguity, whether it concerns personally sensitive revelations of maladjustment or not, and

whether it is likely to be seen as self-descriptive by a certain class of respondents. Although it was a bold stroke to try to finesse the thorny problems involved in the semantics and 'dynamics' of personality item response (Meehl, 1945), this is not an ultimately satisfying solution (Jackson, 1971; Meehl, 1972).

Asking questions is probably the most common method by which clinicians of all kinds gather information regarding their patients or clients. Structured questionnaire items provide controls that are not possible in the unstructured interview. Controlled study of the causal influences affecting the answers respondents give to clinically significant questions is a subject eminently worthy of study. If anything, this is more true today, with the increase of computer-interpreted questionnaires and inventories. However, the response process is quite complex (as Fig. 1 attempted to reflect), and our theory and research method must therefore become complex as well.

REFERENCES

Angleitner A., John O. P. and Loehr F. (1985) It's *what* you ask and *how* you ask it: An itemmetric analysis of personality questionnaires. In *Personality Assessment via Questionnaires: Current Issues in Theory and Measurement* (Edited by Angleitner A. and Wiggins J. S.), pp. 61–108. Springer-Verlag, Berlin.

Block J. (1965) *The Challenge of Response Sets*. Appleton–Century–Crofts, New York.

Bond J. A. (1984) Consistency of test–retest responses to dichotomous personality inventory items: facts and artefacts. Unpublished doctoral dissertation, University of Michigan.

Bond J. A. (1986) Inconsistent responding to repeated MMPI items: is its major cause really carelessness? *J. person. Assess.* **50**, 50–64.

Cliff N. (1977) Further study of cognitive processing models for inventory response. *Appl. psychol. Meas.* **1**, 41–49.

Coppen A. and Metcalfe M. (1965) Effect of a depressive illness on M.P.I. scores. *Br. J. Psychiat.* **111**, 236–239.

Dahlstrom W. G., Welsh G. S. and Dahlstrom L. E. (1972) *An MMPI Handbook, Vol. I: Clinical Interpretation*. University of Minnesota Press, Minneapolis.

Dahlstrom W. G., Welsh G. S. and Dahlstrom L. E. (1975) *An MMPI Handbook, Vol. II: Research Applications*. University of Minnesota Press, Minneapolis.

Edwards A. L. (1953) The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *J. appl. Psychol.* **37**, 90–93.

Edwards A. L. and Walsh J. A. (1963) Relationships between various psychometric properties of personality items. *Educ. psychol. Meas.* **23**, 227–238.

Fekken G. C. and Jackson D. N. (1983) Relative efficacy of four models for predicting the stability of individuals' item responses. Paper presented at the Annual Convention of the Canadian Psychological Association, Winnipeg.

Goldberg L. R. (1963) A model of item ambiguity in personality assessment. *Educ. psychol. Meas.* **23**, 467–492.

Goldberg L. R. (1978) The reliability of reliability: the generality and correlates of intra-individual consistency in responses to structured personality inventories. *Appl. psychol. Meas.* **2**, 269–291.

Hampson S. E. (1982) Person memory: a semantic category model of personality traits. *Br. J. Psychol.* **73**, 1–11.

Harris J. G. and Baxter J. C. (1965) Ambiguity in the MMPI. *J. consult. Psychol.* **29**, 112–118.

Jackson D. N. (1971) The dynamics of structured personality tests: 1971. *Psychol. Rev.* **78**, 229–248.

Kendell R. E. and Di Scipio W. J. (1968) Eysenck Personality Inventory scores of patients with depressive illnesses. *Br. J. Psychiat.* **114**, 767–770.

Kuncel R. B. (1981) Reducing diversity in subject interpretation of items. In *Problems with Language Imprecision* (Edited by Fiske D. W.). Jossey-Bass, San Francisco.

Markus H. (1977) Self-schemata and processing information about the self. *J. person. soc. Psychol.* **35**, 63–78.

Meehl P. E. (1945) The dynamics of 'structured' personality tests. *J. clin. Psychol.* **1**, 296–303.

Meehl P. E. (1972) Reactions, reflections, projections. In *Objective Personality Assessment: Changing Perspectives* (Edited by Butcher J. N.). Academic Press, New York.

Norman W. N. (1967) On estimating psychological relationships: social desirability and self report. *Psychol. Bull.* **67**, 273–293.

Payne F. D. (1974) Relationships between response stability and item endorsement, social desirability, and ambiguity in the MMPI and CPI. *Multivar. behav. Res.* **9**, 127–148.

Rogers T. B. (1973) Toward a definition of the difficulty of a personality item. *Psychol. Rep.* **33**, 159–166.

Rogers T. B. (1974) An analysis of two central stages underlying responding to personality items: the self-referent decision and response selection. *J. res. Person.* **8**, 128–138.

Wiggins J. S. (1966) Substantive dimensions of self-report in the MMPI item pool. *Psychol. Monogr.* **80** (Whole No. 630).

Winer B. J. (1971) *Statistical Principles in Experimental Design* 2nd edn. McGraw-Hill, New York.

APPENDIX

MMPI Group Form booklet numbers of the 83 Neutral items: 1, 4, 12, 25, 60, 70, 77, 78, 81, 83, 87, 90, 92, 95, 98, 101, 113, 115, 116, 120, 126, 132, 140, 144, 149, 150, 164, 173, 188, 195, 196, 203, 204, 207, 219, 221, 223, 225, 249,

255, 256, 258, 261, 272, 276, 283, 285, 295, 300, 370, 372, 376, 378, 423, 428, 429, 432, 434, 435, 440, 441, 445, 446, 460, 463, 483, 493, 497, 508, 513, 514, 529, 537, 538, 546, 550, 552, 554, 557, 561, 562, 563, 566

MMPI Group Form booklet numbers of the 83 Maladjustment items: 8, 41, 52, 57, 61, 67, 76, 79, 84, 86, 88, 91, 94, 99, 104, 106, 122, 138, 142, 158, 171, 172, 180, 201, 202, 207, 209, 210, 217, 244, 259, 264, 267, 292, 304, 305, 309, 321, 337, 338, 339, 357, 361, 371, 374, 375, 377, 379, 382, 384, 389, 390, 391, 395, 396, 397, 398, 407, 411, 413, 414, 416, 418, 431, 449, 450, 453, 455, 479, 482, 487, 502, 509, 517, 518, 520, 521, 526, 531, 543, 547, 549, 555