CASE REPORTS

On the Identification of Warded-Off Mental Contents: An Empirical and Methodological Contribution

Leonard M. Horowitz
Stanford University

Harold Sampson, Elkins S. Siepp, and Alby Wolffson
Mt. Zion Hospital and Medical Center, San Francisco, California

Joseph Wexler
San Francisco Psychoanalytic Institute

The basic assumption behind psychoanalytic psychotherapy is that mental contents that were once warded off come to emerge in the course of a successful treatment. This paper focuses on specific contents which a patient became aware of during a psychoanalytic treatment. Study 1 described the method for identifying warded-off contents: Thems were identified that emerged for the first time between hours 81 and 100; then 20 times analyzed the process notes of the first 80 hours of treatment and judged which of the newly emerging themes had previously been warded off. The judges' ratings were highly reliable and case specific. Studies 2 and 3 validated the ratings (a) by assessing the patient's diaries when warded-off themes emerged into (b) by examining relevant changes in the patient's narratives of early events. Finally, Study 3 explored certain conditions favoring the patient-therapist interaction that facilitate the emergence of warded-off contents.

The basic assumption behind psychoanalytic psychotherapy is that mental contents (cognitions, feelings, impulses) that were once warded off come to emerge in the course of a successful treatment. According to this view, some mental content has assayed or much discomfort in the person that he has to hard himself by wading it off. Because of the delusional process, the context generally remains unacknowledged; however, in the course of a successful therapy, the person becomes conscious of the content and gains control over it (Freud, 1963). Certain theoretical ideas about this process have been discussed recently by Weiss (1975) and Sampson, Weiss, Moosbrugger, and Hauke (1977).

If this view of the therapeutic process is to be systematically studied, we need to be able to identify newly emerging contents that were once warded off. Once this has been done, we could examine the conditions that facilitate their emergence.

This study was supported in part by Grant MH-7144 of the National Institute of Mental Health. The authors would like to thank Cynthia Johanson, Albert K. Levy, and the many trainees and staff members of the Mt. Zion Psychiatric Clinic who generously supplied the data reported in this article. Request for reprints should be sent to Leonard M. Horowitz, Department of Psychology, Stanford University, Stanford, California 94305.

445

445

The basic assumption behind psychoanalytic psychotherapy is that mental contents that were once warded off come to emerge in the course of a successful treatment. This paper focuses on specific contents which a patient became aware of during a psychoanalytic treatment. Study 1 described the method for identifying warded-off contents: Themes were identified that emerged for the first time between hours 81 and 100; then 20 times analyzed the process notes of the first 80 hours of treatment and judged which of the newly emerging themes had previously been warded off. The judges' ratings were highly reliable and case specific. Studies 2 and 3 validated the ratings (a) by assessing the patient's diaries when warded-off themes emerged into (b) by examining relevant changes in the patient's narratives of early events. Finally, Study 3 explored certain conditions favoring the patient-therapist interaction that facilitate the emergence of warded-off contents.
meaning, accuracy, and importance. Also, this research has not demonstrated that the inter- patient contact was previously warded off. Similar remarks have been made by other writers (e.g., Fiske, Hunt, Labovitz, Orne, Parloff, Reiner, & Tuma, 1970; Goodall, 1966; K healer, 1953). The present paper focuses instead on specific contents with a particular patient—contents which the patient became aware of during the therapy and which are shown to have had a dynami- cally significant. This paper has several goals. First, it is designed to describe a method for identifying worded-off contents that emerged in the course of a psychosomatic treatment; this method is described in Study 4. Second, it pro- vides data to show that the method is reliable and valid; Studies 2 and 3 provide overarching lines of evidence which show that the identified contents had in fact been worded off by the pa- tient. Finally, Study 4 examines some circum- stances in a therapy that facilitate the emergence of worded-off contents and relates these findings to a theory of therapy.

Study 4: The Effects of and Its Relevancy

This study was based on a psychosomatic case selected because it was the only tape recorded case that was available to the authors. The pa- tient, treated by psychosomatics for over 3 years, was Mr. B, a young man in his early twenties; he was the third of four children (three brothers and a younger sister). His father was a wealthy businessman and his mother was a housewife. He had recently graduated from college and he came to treatment because he was feeling depressed, lonely, fragile, and sensitive to criticism. He had experienced sexual difficulties (impotence and premature ejaculation) and was currently un- decided about a career.

Because the case was intended as a research case, every session of the analysis was tape recorded with the written consent of the patient. The analyst also wrote process notes at the end of each hour. These notes described material that the patient produced during the session; they did not include only selected, hypotheses, or inferences that might reveal the analyst's thinking.

Case Description

The following information is included for the reader's interest. It was not available to the judge or the authors when the study began, but was needed for executing the studies.

The case can be organized around Mr. B's early relationship with his parents. His mother was a controlling, intrusive, and intimidating woman who would not let him be inde- pendent. He and his family were often at odds, but was unable to reject her because of guilt feelings. When he did yield to her wishes, he felt helpless and unimportant. His relationship over chennes to his mother generalized to other women, and the ambivalence of distance, guilt, and anxiety left him at times excessively dependent. During the first 100 hours of the treatment, he became more aware of his identification with his mother, but it was much later in the treatment that his guilt was reduced enough to allow him to feel superior to and different from his mother. His father, on the other hand, did not protect him from his conflict with his mother or provide him with any directive alternative. He wanted to feel close to his father and win his father's admiration but was unable to do so. Furthermore, he felt that his father favored his two older brothers, who had both been successful in school and in athletics. He tried to acquire a stable identity for himself by assuming the roles of different people (he was usually met and admired, e.g., a writer, a Jewish scholar, a group theo-

Such fantasies temporarily helped him feel stronger, but he then became preoccupied with whether he was real or fake. Furthermore, the distance between himself and his father was so distressing that he came to deny any negative feelings toward his father. Thus, the first 30 sessions showed little improvement toward the father. The references were primarily positive, admiring, and identifying and only later revealed the negative side of the relationship as well.

Mr. B's interaction with the analyst was an important vehicle for this change. During the first 30 sessions, the analyst advised him, direct him, resolve his conflicts, and so on. In this way, he invited the analyst to be the powerful authority and treat him like a helpless child. When the analyst did not fulfill him, however, Mr. B was able to become increasingly stubborn, testy, and uncoop- erative; he came late for appointments, delayed paying his bill, and refused the analyst with in- formation that he was withholding. As described in Study 4, he used these interactions to acquire an increased capacity to criticize, get angry, and disagree with the analyst. At the same time, he began revealing negative feelings toward his father by reciting his father's selfishness and sadism and by acknowledging his own disappointment, frustration, and hostility.
Various changes occurred in Mr. T.'s behavior during the treatment. Initially, he could not involve himself in any job or get close to a woman; he felt crazy, unreal, superficial, and depressed. Eventually, however, he became less depressed, enrolled in an excellent professional school, and academic success there, and developed a deeper relationship with a woman.

Method

Selection of material. First, mental contents had to be identified that emerged in the course of the treatment and might be judged to have been washed off. Two psychologists read the subject's process notes for Hours 1-100, looking for all new themes that emerged between Hours 41 and 100 (i.e., themes that had not appeared in the first 40 hours). They then rated the process notes to check that each theme was actually new. Finally, they compared notes and located 50 themes that they agreed to be new.

Each theme was summarized by a simple statement designed to be rated by judges. (The 50 statements are shown in the Display.) Each statement summarizes a thought, perception, with, impulse, or memory that occurred to the patient about past or present events and referred to the patient himself, his parents, his siblings, his friends, or the analyst.

One theme, for example, which emerged in Hour 65, concerned his father's goodness. It was summarized by the statement, "He thinks of his father as greedy." As a check that judges would not identify previously mentioned themes as washed off, two additional themes were selected from the first 10 hours as dummy items. One theme, "He feels looking ridiculous," emerged in Hour 5, and the other, "He has experienced disgust and pleasure in oral sex," emerged in Hour 10. These themes were also included in the set to make a total of 32 statements.

A simple statement out of context is necessarily ambiguous, because one cannot tell whether the theme emerged by itself or as an incidental part of some broader idea. Therefore, a brief fragment (4 or 5 typewritten lines) was taken from the process notes to show the context in which the idea had emerged. For example, for the statement, "He is like his brother in his orderliness," the passage from the process notes read:

NEWLY EMERGING THEMES (HOURS 41-100)

1. He has never held his brother's baby.
2. He thinks of his father as greedy.
3. He was not as good as much as Ronny and his father.
4. He was afraid that a monster would enter his bedroom.
5. He felt like his mother in his orderliness.
6. It is confused as to whether or not he enjoyed himself last night.
7. He was left in masturbatinal.
8. He has fantasies of leaving people.
9. He wants to show that he is brighter than his therapist.
10. His image of himself seems to be phony.
11. After he saw a girl, he wants to be away from her.
12. He is not adapted to his father.
13. He felt degraded serving drinks at Paul's.
14. He wants the therapist to regard him as special.
15. His father and brothers were less perfunctorial than he.
16. The competed with his brothers for his parents' love.
17. His father was the dominant member of the family.
18. He is a timid small woman.
19. He was envious of his sister.
20. He wanted to assault himself instead of the therapist.
21. He pretended to love his sister.
22. Sometimes he thinks he might be going crazy.
23. He has experienced disgust and pleasure in oral sex.
24. He is convinced of a capacity to enjoy or do things with sex.
25. Sex with girls is a compulsive attempt to prove his masculinity.
26. He always has to have someone to hate.
27. He is imitated by his mother's physical appearance.
28. A green nobby club was important to him.
29. He is jealous of his brother's wife.
30. It would upset him to think that he is irrational.
31. He feels looking ridiculous.
32. He felt humiliated when other people criticized his sex practice.
33. He feels detested being critical of his family.
34. He thought he was mentally ill.
35. It is afraid his dreams of greatness may never be fulfilled.
36. He had a homosexual experience in the sixth grade.
37. He wonders if he is dependent on the analyst.
38. He was interested in studying his mother's mode body.
39. He has marvelous ideas.
40. He needs to be the center of attention.
41. Ronny needs to teach him for his poor self-esteem.
42. When he gets what he wants, he does not want it.
43. He wants the therapist to tell him what to do.
44. His father's death might not completely destroy him.
45. He is angry at his father for indulging him.
46. His mother pleaded him from his father's anger.
47. He does not feel like, cooperating with the therapist.
48. He wanted his father to play baseball with him more.
49. He wanted to dress up for a girl at Halloween.
50. He is afraid he will maul someone if he goes for long.
He thought of his mother’s serene cleanliness and unflawed beauty. She would never trust others to do the dishes and had to do them herself. The patient felt he has taken over her characteristic of being very neat and orderly now that she is living alone.

The judge read the transcript along with the statement, but he was asked to make a judgment about the statement per se. This procedure was designed to maximize both the reliability and the validity of the judgments: The judge rated only a single statement, thereby increasing the reliability, but he also read the accompanying passage to achieve higher validity.

**Jury's ratings.** Twenty clinicians, including psychiatrists, psychologists, and social workers who were advanced trainees and staff members at the Mt. Zion Psychiatric Clinic, were asked to consider each statement and judge whether it had been warded off early in the treatment. Each judge was given a booklet that contained several sections.

First the judge received minimal background information about the patient (e.g., his age and presenting complaints). Then the judge read the analyst’s process notes for the first 10 hours.

Because these notes mainly described what the patient had said, they could not directly cue the judge as to what ideas were warded off, so the judge had to draw his literature from indirect cues (e.g., from a lack of ambivalence where ambivalence is usual).

Then the judge was shown the 52 statements, each with its accompanying passage. He was told that the themes had emerged between Hours 41 and 45 and was asked to decide whether he thought that idea had previously been warded off. He was told to use whatever criterion he wanted, but that he see possibility he want to ask himself whether the patient could have produced that idea in the first 10 hours, or whether he would have aroused too much anxiety for him to face it early in the treatment.1 The

1 Each content emerged for the first time after Hours 41, but not necessarily warded off. Various reasons could exist for not having warded off an idea. For example, the patient might have recognized the idea earlier but simply had no occasion to emote it. Or he might have deliberately suppressed it. Or the idea might simply never have occurred to him before (for nonsymptomatic reasons). A judge would only classify the idea as warded off if he felt that earlier the content would have occurred so much dysfunction that the patient would have had to avoid it through some indirect, active, or passive repressions. No such repressions are implicit in this judgment, and these criteria will be examined later in the general discussion of repressions.

Judge was asked to make a rating on a 5-point scale. A rating of 5 meant that he strongly believed that the idea was not warded off, and a rating of 1 meant that he strongly believed that the idea was not warded off. Intermediate ratings represented varying degrees of uncertainty. The higher the rating, the more strongly the judge believed the idea to have been warded off.

Finally, after the judge completed all of his ratings, he was asked to consider the statements he had rated 4 or 5 and to fill out a checklist for each one. The checklist supplied 11 reasons for classifying a given statement as warded off, and he was asked to check off all of the reasons that applied.

**Analysis of ratings.** The analysis was also asked to rate each statement on a 5-point scale. His ratings were then correlated with those of the judges.

**Control raters' ratings.** A judge’s ratings, in part, could be based on the judge’s a priori ideas about what is generally warded off and what is not. For example, judges might generally consider intense aggressive and sexual impulses to be warded off, and a priori stereotypes of this kind, though not tied to details of the specific case, could produce agreement among judges. In that case, reading 10 hours of process notes would not contribute significantly to a judge’s judgments.

Therefore, another group of clinicians, comparable to the original group of judges, was asked to make ratings of the statements. This group, however, was not given the process notes of the patient, so they had to make their ratings without the case-specific information of the process notes. For that reason these people are called the control raters.

Altogether there were 20 control raters, consisting of psychiatrists, psychologists, and social workers, who were advanced trainees and staff members at the Mt. Zion Psychiatric Clinic. They were asked to rate the 52 statements on the same 5-point scale. Half of them received no information about the patient and half received the minimal background information presented above about the patient. These two subgroups were alike in all important respects. They did not differ in the nature of their ratings or in the degree of correlation of their ratings with either the judges’ or the analyst’s ratings. Therefore, the two groups have been combined in the single report of 20 control raters. Their ratings were averaged and these values were correlated with the corresponding data from the judges and the analyst.
The ratings of the 20 judges were averaged for each statement and the resulting mean was taken as the scale value for that item. These values were averaged across all statements (mean = 3.72; standard deviation = 0.85). The lowest value was 2.00 for the item, "He is looking ridiculous." and the highest value was 4.50 for the item, "He thinks of his father as greedy." The split-half reliability for the 52 items was computed by correlating the mean values of one randomly selected subgroup of 10 judges with the corresponding mean values of the remaining subgroup of 10 judges. The resulting split-half reliability coefficient was 85. Finally, as expected, the two dummy items that were chosen from the early hours correctly received the lowest two values of the similar set, 1.50 and 1.75.

The analyst's ratings ranged from 1 to 1. When these values were averaged, they yielded a mean of 3.12 and a standard deviation of 1.30. The analyst's ratings were correlated with the values obtained from the panel of judges to produce an r of .80, significant at p < .001.

Table 1 shows the 12 items that received the highest judges' ratings and the 12 that received the lowest ratings. The analyst's ratings for these items are also reported. Notice that the analyst had assigned a rating of 1 to each of the top statements and he had generally assigned a rating of 1 (occasionally 2) to the bottom statements. These features further reflect the close agreement between the judges and the analyst.

Finally, the mean ratings of the control items were examined. The resulting values ranged from 3.05 to 4.0. The overall mean was 3.11, with a standard deviation of .61, and the split-half reliability was .83. However, the control rates' ratings did not correlate significantly with either the judges' ratings or the analyst's ratings. The former r was .50, the latter r, .68, with r > .10 in both cases.

**TABLE 1**

<table>
<thead>
<tr>
<th>Statement</th>
<th>Mean of judges' ratings</th>
<th>Mean of analyst's ratings</th>
<th>Mean of control judges' ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>He thinks of his father as greedy.</td>
<td>4.50</td>
<td>5</td>
<td>3.75</td>
</tr>
<tr>
<td>He left jealousy of his sister.</td>
<td>4.20</td>
<td>5</td>
<td>3.15</td>
</tr>
<tr>
<td>His evil at his father for instilling in him.</td>
<td>4.05</td>
<td>5</td>
<td>3.05</td>
</tr>
<tr>
<td>He denied his father.</td>
<td>4.15</td>
<td>5</td>
<td>2.80</td>
</tr>
<tr>
<td>He wants to show that he is brighter than his stepmother.</td>
<td>4.00</td>
<td>5</td>
<td>3.85</td>
</tr>
<tr>
<td>His father's death might not completely destroy him.</td>
<td>4.00</td>
<td>5</td>
<td>3.35</td>
</tr>
<tr>
<td>He has intimates of bathing people.</td>
<td>3.95</td>
<td>5</td>
<td>3.05</td>
</tr>
<tr>
<td>He went to dinner with a girl on Halloween.</td>
<td>3.00</td>
<td>5</td>
<td>4.00</td>
</tr>
<tr>
<td>He does not have someone to hate.</td>
<td>3.85</td>
<td>5</td>
<td>3.60</td>
</tr>
<tr>
<td>He has numerous lies.</td>
<td>3.80</td>
<td>5</td>
<td>3.75</td>
</tr>
<tr>
<td>His mother protected him from his father's wrath.</td>
<td>3.80</td>
<td>5</td>
<td>4.25</td>
</tr>
</tbody>
</table>

**Lowest ratings:**

- He looks ridiculous.
- He has experienced disgust and pleasure in oral sex.
- After he scores a goal, he wants to be away from her.
- He is compared as to whether or not he enjoyed himself the night before.
- He diagnoses the therapist for not being more realistic toward women.
- He wants the therapist to tell him what to do.
- He needs to be the center of attention.
- When he knows what he wants, he does not want it.
- He is expected to live up to his physical appearance.
- He does not feel like cooperating with the therapist.
- He was in love.
These results suggest that the informed judges, like the analyst, used case-specific information derived from the process notes; their judgments are not based simply on a priori notions of what is warded off. In fact, marked discrepancies exist between the control rats' ratings and those of the judges. The item with the highest rating, an average of 4.40, by the control rats was "He was seen eating in his mother's nude body." informed judges, however, rated the item somewhat lower at 3.60. Table 1 shows other discrepancies of this kind.

As mentioned above, the judge was also asked to consider each item that he had rated 4 or 5 and to check all the reasons that existed for labeling that item as warded off. The mean number of reasons checked was 2.17 for statements rated 4, and 4.16 for statements rated 5. These means differed significantly, F(1) = 3.35, p < .05. Thus, the higher rating was used when more reasons existed for classifying the item as warded off.

One of the reasons on the checklist concerned the nature of the content itself: "The content of the statement per se is suggested that the idea had been warded off." Another reason referred to the judge's own formulation of the case: "My case formulation suggested that the content had been warded off." (A case formulation is a theory about the case that a judge develops to be able to derive case-specific clinical inferences. Other reasons on the checklist were more general, for example, "My subjective experience (introspection, clinical intuition, or empathy with the patient) suggested that the content had been warded off." or "My clinical experience with another case similar to this one suggested that the content had been warded off."

The number of times a judge checked a particular reason was recorded and expressed as a proportion of the number of opportunities for checking that reason. This proportion was computed separately for every judge for the items rated 4 and for the items rated 5. These proportions were then averaged across judges. The reasons checked most often were those concerning the case formulation. The most proportion was .83 for statements rated 4, and .49 for statements rated 5. The second most frequently checked reason was the one concerning the judge's subjective experience; the means were .40 and .42. Other reasons were checked substantially less often, usually less than .15 of the time. According to the judges' reports, then, they frequently developed and used a case formulation as a basis for their ratings.

To summarize, judges made reliable judgments largely by formulating the case from the process notes of the first 10 hours. They agreed well with the analyst, but they did not agree with naive control rats who had no way of formulating the case. The ratings do provide a clue to the nature of the judges' case formulations and suggest agreement among judges in their case formulation.

STUDY 2

VALIDITY OF RATINGS: DISCOMFORT QUOTIENT

Having established the reliability of the judges' ratings in Study 1, our next step was to establish their validity. One line of validating evidence comes from the data of Study 1. A critical examination of those data will provide the rationale for Study 2.

After Study 1 was completed, the analyst was asked to write a brief description of the case. His description stressed that the patient initially idealized the father and was warded off critical ides and hostile feelings toward him. Among the statements of Study 1, nine mentioned the father, seven of them received judges' ratings of 3.5 or higher, and four of those had ratings of 4.0 or higher. (Five statements that expressed nonidealized feelings toward the father are shown in Table 1 among the statements with the highest ratings. For those statements, the mean judges' rating was 4.13, and the mean of the control rats' ratings was 2.99.) Thus, judges detected this tendency in the patient from the limited information given in the process notes on 10 hours. This agreement with the analyst's view of the case reflected, of course, by the p of .80 between the judges' and the analyst's ratings.

Agreement of this kind does not really validate the ratings, however, because the analyst's case formulation can be viewed as simply one additional judge's ratings. One could argue that all of the information of Study 1 concerns only judges' ratings. Although the judges and the analyst are in agreement with thejudges in rating statements necessarily making invalid inferences: they may simply be making correlated errors. Thus, a critic could argue that the validity of the ratings can only be assessed through primary data supplied by the patient himself. Therefore, further steps were needed to validate the judges' ratings.

If the judges' ratings were valid, one might expect the patient to exhibit heightened discomfort at that
moment may not be a necessary condition for its having beenwards off, but the presence of a
discom for would clearly document its troublesome
nature for the patient. Therefore, some measure of the
patient's discom for was needed. Several
investigators (Owen, 1966; Kazal & Malb, 1975;
Malb, 1976, 1979) have shown that discom for
of a person's speech reflect his discom for at the
time the speech was created, and it was
decided to modify Malb's measure (e.g.,
Kazal & Malb, 1975) to derive an index of the
patient's discom for at the time the word of
material emerged.

Method
The six items with the highest judges' ratings
and the six with the lowest ratings (not counting
the two dummy items) were considered. Those
with the highest ratings were called W items (for
"worded off") and those with the lowest ratings
were called N items (for "nonworded off"). Each
item came from a different hour and the 12
hours were typed verbatim from the tape record-
ings of the hours. All hours were typed by the
same typist, and each hour was processed by the
first author. The typewritten included all pauses,
speech dysfluencies, word repetitions, etc.

There was one W item for which the worded
of theme was exceedingly brief, occurring only
for lines of type (less than 20 sec of speech).
Likewise, one N item was also exceedingly
brief. In all the remaining hours—five W hours
and five N hours—the themes listed more than
1 min and was over 14 lines of type. The fol-
lowing procedure was applied to these 10 hours:
Each hour was divided into successive episodes
of patient talk. An episode is defined as an
uninterrupted segment of patient talk not con-
necting or more or less identified and agreed
on each change of theme. An hour always started with the patient talking. The episode ended when the theme changed or when
the therapist interrupted with some comment.
The next episode began with the next period of
patient speech.

On the average, an episode was 20-25 lines of
speech and lasted about 2.5-3 min. Table 2 shows
the five W themes and the five N themes. For
each statement, the table shows the hour in
which the theme emerged, the position of the
episode in the hour, the length of this episode,
and the total number of episodes in the hour.
Two hours have fewer episodes than usual be-
cause the patient arrived late.

The measure which was devised to determine
the patient's level of discom for included four
classes of indicators; a point was scored every
time one of these indicators was noted. The class
included filler words and phrases like, "you
know," "I mean," "I don't know," "gives me,"
and "yeah." A second class contained various
forms of speech dysfluencies: repetitions of a
word, false starts, stammers, or slips of the
tongue. (A repetition was scored whenever the
patient repeated a word or phrase. A false start
was scored whenever the patient began to utter
a sentence but shifted to a new idea before the
first one was completed.) A third class contained
particular sounds: a sniff, laugh, cough, sigh,
'vawn, swallow, or clearing of the throat. Finally,
a miscellaneous category was scored whenever
the patient's speech was unintelligible, whenever
he was gaited, and whenever he erected his
knuckles or clicked some nails.

The total number of instances in an episode
was counted. The total was then divided by the
number of lines of type in that episode to yield
a discom for quotient.

Results and Discussion
The discom for quotient (DQ) was measured
for every episode of each of the 10 hours, yield-
ing values from .38 to 2.67. The overall mean
was 1.33 and the standard deviation was .97.
Reliability, measured by correlating corresponding values obtained by two different scorers, was .97.
Table 2 shows the DQ of the episodes that
contained the five W themes and those that con-
ained five N themes, as well as the mean DQ
for each hour. First note the results for the N
hours. The DQ when the theme emerged was 1.10
on the average, and that for the entire hour was
1.21 on the average. Thus, the episode with the
newly emerging N theme had about the same
level of discom for, whether the hour was
short or not.

The W hours, on the other hand, exhibited
significantly higher DQs. The corresponding
results were substantially higher at 1.33 and 1.51,
and the various DQs showed almost no overlap
with those of the N hours. When an analysis of
variance was performed on the data of the last
two columns in Table 2, the overall difference
between the W values and the N values was
found to be highly significant, F(1, 8) = 35.14,
p < .01. However, neither the column differences
nor the interaction was significant.
That clearly suggests that a newly emerg-
ing W theme shows significantly more discom-
for than does in N theme, also, the W hours in
general show significantly more discom for than
the N hours. Accordingly, an hour in which a once
worded-off content emerges generally contains
more discomfort, both at the time the content occurs and elsewhere throughout the hour. This point, in part, validates the judges' ratings. The discomfort quotient, which was independent of the judge's ratings, objectively measured the patient's discomfort and thereby differentiated the U and N themes. Thus, from their limited information, the judges validated theories that could independently be shown to have aroused discomfort in the patients.

STUDY 3

VALIDITY OF RATINGS: CUMULATIVE MEMORIES

As mentioned above, a number of items with high ratings were statements that criticized the father. The ratings implied that initially the patient could not tolerate these critical feelings, though later he became to do so. Thus, the judges' ratings implied a change in the patient's manifest attitude toward the father from initially positive to negative. This attitudinal shift should be directly observable in the patient's references to the father.

Not all references to the father are useful, however. Changes observed in the patient's description of his current relationship with the father, for example, might be due to a real change in the father-son relationship. However, if we are to show that critical feelings had been washed off, the newly emerging feelings must be attached to events that preceded the therapy. Therefore, memories of the father were examined, all of which reported relatively early events.

Study 3 therefore examined one class of situations, namely, memory about the father. Two groups of memories were examined: those that occurred relatively early in the treatment and those that occurred later. The hypothesis claimed that an attitudinal change would be observed as the patient reported memories of the father that were initially positive and idealized but later negative and critical. This study formed part of a larger methodological project, which will be reported elsewhere, concerning the sensibility of
process notes in detecting significant therapeutic changes.

Method
Two blocks of sessions were examined: Hours 1–20 and Hours 60–90. Equipment difficulties arose during Hour 60, so none of that hour was not available. A listener, who was not familiar with the goals of this research, listened to all of these tape-recorded hours and isolated all memories that mentioned the father. These memories were then typed verbatim from the tapes and placed in random order.

Two judges were asked to sort the memories into two categories that were termed positive and negative memories. A positive memory was one that (a) expressed admiration for the father (e.g., how masculine and strong he was), (b) showed how dependent the patient was on him, or (c) expressed a satisfaction with closeness with him. A negative memory was one that (d) described the father as not admirable (e.g., how he sometimes became cruel and distanced), (b) showed a conflict between him and the patient, or (c) expressed a frustration with closeness with him. Explicit scoring rules were devised to specify these criteria in detail. The two judges used the rules to sort the memories into the two categories.

Results and Discussion
Altogether 37 memories were isolated—or 15 in the early block and 22 in the later block. The two judges agreed on their classification in 95% of the cases; they later compared to resolve the disagreements.

Following is an example of a positive memory from Hour 19:

This picture reminds me... a whole lot... of our country place, you know. I met the feeling... of the same kind of warmth that I was... because... in the... one... you know, it's kind of a thing where you know. I feel like I'm walking through a field with my dad, you know. When I see this and that later's go back to the house and we'll sit down and we'll have lunch and a big fire. Jews, I miss that so much.

An example of a negative memory is the following from Hour 79:

These were certain times when my father... and... when he'd really get mad and hit us. There are famous automobile accidents when my dad's experience for hitting us was "bashful" as, when he would warn us children to the best seat, you know, to shut up and quit playing around and making so much noise... and the waddlings would really cause... "Oh, I'm going to pull over the car and bash you," and if it didn't stop, sometimes he just really did it, you know. Sometimes in anger, you know, you could see his... the marks in his... tightening so... and he would just lose it. He would pull over and really break in. I was always terrified at him really doing that. I was always ter-

There were 15 memories in the early block of hours; 22 memories were classified as positive and 5 were classified as negative. On the other hand, there were 22 memories in the later block of hours, 9 of which were classified as positive and 13 as negative, $x^2(1) = 4.11, p < .05$ (with a correction for continuity).

Additional memoires of the father occurred in a still later block of eight hours (by Hour 200) that had to be examined as part of the larger study mentioned below. This block contained one positive and seven negative memories of the father. By adding these results to those reported above, a 3 X 5 contingency table was prepared, $x^2(2) = 10.65, p < .005$. Because of the slow expected frequencies in some cells, the two later blocks of hours were combined (Hours 61 and beyond), $x^2(1) = 8.33, p < .05$ (with a correction for continuity). These results document the shift in the patient's attitude toward his father as the therapy progressed. The shift was implied by the judges' ratings, so it adds further evidence for the validity of the ratings.

These results can also be taken as empirical evidence for an early defensive attitude which later gives way to a more negative attitude. This kind of shift has been the subject of a recent paper by Suppe and Winer concerning self-attitudes. The authors have proposed that a defense mechanism be formally viewed as a transformation on a progression. The original proposition represents the unconscious idea that eventually emerges, whereas the transformed proposition represents the original, conscious idea. The marked shift in this paper readily allows an investigator to identify a mental content that has undergone such a shift and to systematically study the frequency and nature of such shifts.

STUDY 4

Conditions Affecting the Emergence of Warmed-Off Content

The following study explored certain conditions that facilitate the emergence of W content. It was prompted by an evolving theory (Simpson et al., 1972; Weiss, 1971) which will be described more fully in future publication, but which can be partly summarized as follows:...
A patient in therapy wants to bring forth and master warded-off content. To accomplish this goal, he first needs to assure himself that it is safe to do so. Safety, in part, depends on the therapist's reaction at critical times, as during therapy the patient continually tests the therapist and evaluates his reactions.

These tests and their consequences are significant to the therapeutic process. To pass a test that the patient poses, the therapist should remain neutral. For example, one way for the therapist to be neutral is to objectively investigate issues underlying a conflict without taking a stand that favors either side of the conflict. For example, the therapist should not condemn or criticize the patient, placate him, or act sad, angry, or guilty. A therapist's response of the latter type would be nonneutral. When the therapist is neutral, however, the patient feels safer, and when he feels sufficiently safe, he lifts his defenses and brings forth warded-off content. Thus, as one major sequence in a therapy, a test is performed, the therapist passes the test by remaining neutral, and the patient feels safer; then the patient lifts his defenses and formerly warded-off content emerges.

Various kinds of tests help the patient assess the situation's safety. For example, one test occurs when the patient disagrees with the therapist, another when the patient is angry at the therapist, and a third when the patient makes a demand of the therapist. The therapist's failing or passing the test, of course, depends on how the patient diagnostically views the therapist's response. However, in many psychoanalytic interactions, judges would probably agree about the neutrality or nonneutrality of the therapist's response to a disagreement, an expression of anger, or a demand. In the following study, 23 such interactions were identified and 3 judges assessed the degree of neutrality of the therapist's response. Tests that the therapist had passed were then compared with those he had failed.

When a therapist responds neutrally, and hence passes the test, two consequences should follow. First, the patient should tend to show immediate signs of relief; the DQ should generally drop when the test is passed. Second, feeling safer, the patient should be more apt to bring forth warded-off content; therefore, the W hours should contain significantly more passed tests.

Method

Two psychologists read the 12 W and 11 hours, looking for all episodes in which the patient (s) openly disagreed with the analyst, (e) expressed anger at the analyst, or (c) made a demand of the analyst. The psychologists agreed perfectly in identifying episodes: 14 episodes occurred in W hours and 9 occurred in N hours. (Every W hour had at least one of these episodes, as did four of the N hours.) Altogether, there were 8 cases of open disagreement, 5 cases of anger, and 10 cases of demands.

First, the episodes (as defined in Study 2) were identified in which these incidents occurred; these episodes are called the "critical episodes." The analyst's response was also noted, as well as the episode that followed it (the "following episode"). The discomfort quotient was measured for each critical and each following episode, as noted above, the DQ had an interjudge reliability of .90.

In addition, each critical episode was typed on a page together with the analyst's response. Three clinical psychologists were asked to judge how neutral the analyst's response had been. If the therapist had made no mistake in his neutral response, his response was considered neutral.

Results and Discussion

The 23 critical episodes can be divided into two major groups. First consider the 10 cases in which the therapist has not passed the test in which he had failed. When a therapist responds neutrally, and hence passes the test, two consequences should follow. First, the patient should tend to show immediate signs of relief; the DQ should generally drop when the test is passed. Second, feeling safer, the patient should be more apt to bring forth warded-off content; therefore, the W hours should contain significantly more passed tests. Study 4 examines these two hypotheses with a method resembling that of the symptom-content method (Luborsky, 1967, 1970; Luborsky & Auerbach, 1969).

As an example of an open disagreement, the following excerpt from Hour 44: "I don't see that. Whenever I admit distortions in the way I think about myself... I like to think that my judgments are accurate, that they represent the way things are. I don't want to make them come out as they are. I think that this kind of statement is like a way of saying I don't want you to know, you know, point me in some direction."

This description only occurs on one strand of the therapeutic process. It does not consider other important ingredients of a therapy that facilitate the emergence of warded-off content, such as the therapist's intervention, indeed, when expressed, how imply that once-warded-off contents can sometime emerge without any explicit therapist interpretation for the therapist. A fuller treatment of this issue, along with empirical data, will be the subject of future investigations.
which the analyst said nothing. Seven of these cases occurred in W hours, 3 occurred in N hours. The DQ was measured for each critical episode and for each following episode. In every case, the DQ of the following episode was lower. The mean DQ for the critical episodes was 1.53 and for the following episodes was 1.26. The difference was significant, \( t(9) = 14.90, p < .001 \). Thus, the DQ dropped when the analyst said nothing.

Part of this effect is due to the patient's generally greater discomfort during a critical episode. In seven cases, the patient's DQ was higher than the mean DQ for that hour. Regression toward the mean would therefore yield some average reduction in DQ. Two facts, however, show that the drop in DQ exceeded the amount expected by regression. First, episodes with a DQ below the hour's mean also showed the drop; in three cases, the DQ of the critical episode was below the hour's mean and each of those cases showed the drop as well. Second, the DQ of the following episode often fell below the hour's mean; that result occurred in six cases and would not be explained by regression toward the mean.

Now consider the 13 cases in which the therapist did make some comment. These comments varied in rectrularity, and the three judges ranked ordered them for their neutrality. The judge was also given a checklist showing different ways in which a comment might depart from neutrality. For example, the analyst might have insulted the patient, disagreed with him, seemed anxious, or seemed defensive. The judges were asked to check all of the ways in which the comment departed from neutrality.

The judges agreed well in ranking the fragments; the coefficient of concordance (Kendall, 1955) was .71. The ranks were averaged and the 13 patients were divided into 2 subgroups: the 8 fragments that seemed most neutral (the judges agreed perfectly in their choice of the top 6) versus the 7 that seemed most non-neutral. For each case, the DQ was noted for both the critical and the following episodes; the change score was also noted. For the more neutral fragments, the mean change was -49 (an average increase); the DQ increased in six out of seven cases. Fisher's exact test was used to compare these frequencies for the neutral and non-neutral fragments. The difference was significant at \( p = .025 \).

Thus, cases that showed a drop in DQ were generally those in which the analyst had been neutral. The data further showed that cases of this kind (reduced discomfort after a neutral intervention) occurred more often (four out of five times) in the W hours. On the other hand, cases that showed a rise in DQ were generally cases in which the analyst had been non-neutral. Cases of this kind (increased discomfort after a non-neutral intervention) occurred more often (five out of six times) in the N hours. The chance probability of this result is .007 by the Fisher exact test.

If the data of all 23 instances are combined, there were 16 cases in which the analyst was neutral and 2 cases in which he was non-neutral. Of the neutral category, 15 showed a drop in DQ after the intervention and 1 showed a rise. Of the non-neutral category, 1 showed a deep in DQ and 6 showed a rise. The 15 episodes showing decreased discomfort after a neutral intervention mainly occurred in W hours; 11 were in W hours and 4 were in N hours. The 6 cases showing increased discomfort after a non-neutral intervention mainly occurred in N hours; 5 occurred in N hours and 1 occurred in a W hour. This difference was significant by the Fisher exact test at \( p = .029 \).

These results suggest that the patient's discomfort drops when the therapist is neutral. A patient exhibits more discomfort when he is testing the therapist, but the therapist's neutrality seems to reduce the discomfort. Furthermore, W hours contained more sequences in which (a) the patient posed a test, (b) the therapist responded neutrally, and (c) the patient's DQ dropped.

It is also worth noting that the W content generally emerged after a test was passed. Of the 14 tests that occurred in the W hours, 10 preceded and 2 followed the emergence of the W content. Thus, the patient seems to have performed the test to determine whether the therapeutic situation was safe; when the test was passed, he then brought forth the W theme. On the other hand, W themes did not depend on the test. The N theme preceded or followed the test about equally often of the nine tests that occurred in the N hours; four preceded and five followed the emergence of an N content. The chance probability of a pattern his extreme is .007 by the Fisher exact test.

Although a definitive statement of causality cannot be made, the results suggest that the therapist's neutrality enhances the patient's trust; the patient then relaxes his defenses and allows the warded-off content to emerge. It is also interesting to note that tests were passed (and W themes emerged) more often in
the relatively early hours—Hours 42, 47, 58, 63, 71, and 76. Of these, on the other hand, seemed to be more productive (as productively defined in terms of the emergence of ward-off tendencies). Indeed, the latter stage seemed to be a new way to characterize the productivity of a block of hours, namely, to correlate the mean ratings of each theme with the hour number. In this ideal case, the two should be positively correlated, as the hours, (progressively), increasingly difficult themes (with higher ratings) should emerge. When the correlation is negative, the therapeutic work seems to be slowing down.

In the present case, the two between the judgment ratings and the hour number for the 50 state-ments (counting the two dummy forms of earlier hours) was —.20. This value differed significantly from 0 and suggests that the therapy was not following an ideal course as it approached Hour 100. In the later hours, the patient's ratings (both did not correlate) positively, so ward-off contents temporarily diminished. Other therapists have also developed measures of therapeu-tic productivity. Daitre (1972) has developed a method based on the factor analysis of emerging themes in a psychodrama. Syme (1972) used clients' ratings of productivity and evaluated correlated characteristics of the patient's work. These methods differ in methodological ways from that of the present paper, but their validity and differences cannot be meaningfully evaluated until all of the methods have been applied to a variety of different cases.

GENERAL DISCUSSION

A mental content is said to be ward-off when it emerges enough in the patient that he avoids it through some defensive activity. In the present analysis, evidence existed to support all three parts of this definition. Thus forming a very stringent set of criteria: First, the content was related to the patient early in the treatment and became more available in later hours. That is, contents were selected for Study 3 that had never occurred in the first 40 hours. Second, the content produced discomfort when it first emerged. Study 2 showed that when the content entered the W themes his speech showed increased signs of discomfort. Third, Study 3 provided objective evidence for a defensive process early in the treatment. One form of evidence that a person is defending against some idea would be a pair of contradictory attitudes or beliefs in which one partly negates the other, thereby equilibrating an anxious state. The shift described in Study 3 provided evidence of this kind.

Although these three criteria are not neces-sarily essential conditions for calling a content ward-off, together they provide a more compelling test of the validity of the judges' ratings than would one or two of the criteria alone. For example, on the one hand, that a content emerged for the first time and caused anxiety yet showed no evidence of having triggered a defensive process. We could not be sure that the content had truly been ward-off if it might have remained unobservable for nonanalytic reasons. (2) Perhaps the patient had not gotten around to examining the idea earlier. (3) Perhaps the idea had not seemed important earlier so the patient had not attended to it, or (4) perhaps other ideas had directed the patient's thoughts elsewhere. Content's of this kind, even those that aroused anxiety, are not ward-off. However, evidence of a derivitive process supports this one important source of ambiguity. On the other hand, suppose the second criterion (the presence of anxiety) were not met. Suppose a newly emerging idea replaced an earlier contradictory idea and neither aroused anxiety. Suppose, for example, that a patient who had idealized his father came to view his father in critical terms, but without anxiety. Then we could not be sure that the anxiety factors had kept the new content from becoming observable. This idea might have been ward-off, but it is also possible that the patient had simply changed his mind about his father. Too little, suppose the W contents of the present study had not aroused anxiety; then we could only say that the judge had anticipated a change in the patient's attitude toward his father, and this attitude change might be explained in other ways. For example, through some subtle means, the analyst may have per-suaded the patient to feel better about his father, and the patient, eager to please the ana-lyst, complied. However, this kind of explanation would not account for the patient's anxiety because a patient who had simply changed his mind to please the analyst should not be made anxious by the new attitude. Thus, the presence of anxi-ety supports such objections and removes a sec-ond kind of ambiguity. It is possible, of course, that a ward-off content might emerge without discomfort. Per-haps a content that was once anxiety producing be-comes more tolerable through implicit therapeutic
mechanisms and then emerged without discom- 
fort. For example, a content might be measured 
outside of the formal therapeutic setting, or 
work on other contents might generate to the 
as yet unverifiable content. Thus, anxiety at the 
time the content emerges is probably not an 
essential condition for the client's having been 
warded off. (For that matter, it is possible that 
no single criterion is absolutely essential. The 
defense mechanism of isolation, for example, 
might allow a content to be present without 
anxiety early in the treatment.) However, when 
an investigator can meet all three criteria, im-
portant sources of ambiguity are removed and the 
result is strengthened for calling the content 
warded off.

Furthermore, the discomfort produced by the 
W contents in Study 1 was a very impressive 
testimony to the judges' skill. These judges, with 
only the limited, anecdotal information of the 
previous notes, were able to detect contents which 
indeed were found to make the patient anxious. 
What mechanism shows a judge to anticipate 
that the W themes would make the patient anx-
ious? According to the data of Study 1, the 
judges relied heavily on their case formulation. 
A case formulation seems to be a prerequisite 
for successful clinician prediction. Among other 
things, it characterizes patient confides that 
pattern consistent cognitive and behavioral) 
stills. That is, the clinician looks for patterns 
that consistently evoke in different situations in 
the person's life (including the kind of inter-
action with the therapist described in Study 4).

In a recent review of the literature, Born and 
Allen (115-11) have stressed the importance of 
cross-sectional trait consistency as a measurable 
individual difference that significantly reap-
porters the prediction of behavior. In their study, 
ratings on a trait's consistency were computed 
along with ratios of the trait itself. Born and 
Allen showed that these two properties need not 
be correlated; the variability of consistency-
ness across situations, for example, is not neces-
sarily related to the person's mean position 
along trait dimension. Their study showed a sub-
stantial increase in the predictability of the sub-
ject's behavior when consistency was included in 
as predictor. 

A case formulation likewise takes account of a 
trait's consistency. When a clinician formulates a 
case, he does not simply cite characteristic 
traits; rather he seeks traits with observed 
cross-dimensional consistency. In the light of each 
session's additional data, the clinician consis-
tently revises early hypotheses to incorporate the 
most consistent patterns that equally emerge. 
The patient of the present case, for example, 
initially showed the same indelible attitude 
toward his father, his therapist, and other au-
thority figures. Using this consistency in one par-
dicator, the clinician (or judge) could then make 
a successful inference. 

To summarize, the present studies provide a 
method for reliably and validity identifying 
warded-off contents. Using this method, we can 
examine cognitive and emotional changes that 
occur within a patient during therapy as well as 
treatment by which judged a clinical in-
ference. The method also shows us to pinpoint 
consistency of the patient-therapist interaction 
and thus follow in detail the fine grain of the 
therapeutic process.

REFERENCES

Zaus, D. J., & Allen, A. On predicting some of 
the patient's state of the: The thrust for cross-
situational consistency in behavior. Psychol-
ological Reviews, 1974, 81, 324-326.

Dahl, M. A quantitative study of a psychoanalyst, 
In R. E. Hest & E. Miller (Eds.), Psycho-
analysis and contemporary science (Vol. 1). 

Dillen, A. S. Case-counting: A means of dealing 
with intercourse, Journal of Consulting Psychol-. 

Fink, D. W., Hare, H. F., Ledovsky, L., Green, M. 
T., Fulford, M., Reber, M. T., & Tunj, A. H. 
Application of research on efficiency of psycho-
therapy. Arch of General Psychiatry, 1975, 
22, 22-25.

Fred, S. On psychotherapy. In P. Roff (Ed.), 
Therapy and effectiveness: The collected papers 
of Napsword Floyd, New York: Collier Books, 
1963.

Gersch, E., & Huggard, T. Immediate effects in 
process of psychoanalytic interpretation. Psychol-

Gendlin, E. E. The social significance of the 
relationship In C. B. Rogers, E. T. Gendlin, 
D. J. Kirschen, & S. B. Penrose (Eds.), The psycho-
analytic movement and its Impact: A study of psycho-
therapy with schizophrenics. Madison: University 

Kend, S. V., & Mah, J. F. Experientially induced 
and social distortion. American Psych-

Kend, S. V., & Mah, J. F. Experimentation and insti-
ution in schizophrenia. Journal of Personality and Social 
Psychology, 1965, 2, 121-130.

Kendall, M. G. Rank correlation methods (2nd ed.). 

Kiesler, D. J. Some aspects of psychotherapy re-
search and the search for a paradigm. Psychol-

Lambert, J. 'Observatory foraging during psycho-
therapy and psychosynthesis: A theory and re-


(Received October 3, 1971; revision received January 7, 1972)