1-1-1996

Splitting and self-schemata.

Joseph A. Bousquet
University of Massachusetts Amherst

SPLITTING AND SELF-SCHEMATA

A Dissertation Presented

by

JOSEPH A. BOUSQUET

Submitted to the Graduate School of the
University of Massachusetts in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 1996

Department of Psychology

SPLITTING AND SELF-SCHEMATA

A Dissertation Presented

by

JOSEPH A. BOUSQUET

Approved as to style and content by:

Sheldon Cashdan, Chair

Richard Halgin, Member

Ronnie Janoff-Bulman, Member

Hannah Kliger, Member

Melinda Novak, Department Head
Department of Psychology

ACKNOWLEDGMENTS

Dr. Sheldon Cashdan's book on object relations drew me to UMASS in 1991 and he was a mentor and clinical supervisor before chairing my committee. My graduation comes a semester after Shelly's retirement from academia and I am proud to be his final doctoral candidate. He has treated me with unfailing decency, good humor, patience, and intelligence.

It has also been a pleasure and privilege to work with Dr. Richard Halgin, Dr. Ronnie Janoff-Bulman, and Dr. Hannah Kliger, the other committee members. As the representative from outside the department, Dr. Kliger's insightful observations and suggestions were particularly noteworthy. I am grateful to Rich Halgin for all he taught me about psychological assessment, as well as his generous service on the committee. Ronnie Janoff-Bulman encouraged me to continue this project at a moment when I was indecisive by sharing some of her thoughts on splitting; I was also lucky to take an excellent course in social psychology from her.

The studies described in this dissertation would have been impossible without the assistance of Sarah Amitay, Nurit Ben-Harazz, Joanna Farewell, Alison McVay, and Alicia Straus. I would like to thank Dr. Lucille Halgin for coordinating the research assistant program. Dr. Paula Pietromonaco, Mary Beth Wade and Beth Morley managed the prescreening process. Areyeh Shestopal, a fellow graduate

student, made many helpful suggestions, and added the Borderline Scale to the prescreening instrument.

Chris Christian, Michelle Jacobo, Julie Kurcias, Nahid Karamali Markosian, Robert Murphy, and Charles Powers, comprised my cohort in the clinical program. They assisted my progress in too many ways for me to do them justice in a brief acknowledgment. Ernie Yonkovitz is another good friend and fellow traveller in the field whose support has been invaluable to me.

I feel privileged to have taken first year statistics from Dr. Jerry Myers, whose enthusiastic style of teaching made certain esoteric bits of statistical wisdom comprehensible. A challenging course in Social Cognition taught by Dr. Susan Fiske inspired certain aspects of the present research. I greatly value what I learned from her. Dr. Drew Weston, of Cambridge Hospital, made some helpful suggestions regarding the second experiment. I would also like to thank Dr. Mort Harmatz for his first-rate clinical supervision, Dr. Pat Wisocki for her general attentiveness to student needs, and Dr. Charles Clifton for refusing to turn in strikers' names during a unionizing effort.

Finally, I wish to thank my mother, (a union stalwart, herself) who was always willing to spend scarce dollars on books and schooling when I was young, and my soulmate and spouse, Gyda DiCosola, for all sorts of sacrifices endured on my behalf.

ABSTRACT

SPLITTING AND SELF-SCHEMATA

MAY 1996

JOSEPH A. BOUSQUET, B.S., FLORIDA STATE UNIVERSITY

M.A., UNIVERSITY OF FLORIDA

M.S.W., UNIVERSITY OF CONNECTICUT

Ph.D., UNIVERSITY OF MASSACHUSETTS

Directed by: Professor Sheldon Cashdan

Two studies test the idea that the psychoanalytic defense of splitting involves alternating activation of global, opposing self-schemata, including a *good-self* schema, and a *bad-self* schema. Subjects were chosen on the basis of scoring very low or very high on Sharon Gerson's (1984) Splitting Scale, allowing frequent splitters (FS) and infrequent splitters (IS) to be contrasted on a variety of experimental measures. Subjects were asked to rate their schematicism on a variety of trait dimensions and to rate two characters in a Thematic Apperception Test (TAT) card on the same dimensions. They were also asked to judge whether perceptibly and subliminally presented trait adjectives were best characterized as *me* or *not me*.

Results indicate that IS identify themselves more schematically than FS on the trait dimensions *good*, *bad*, *loving*, and *hateful*. FS showed longer average reaction latencies in the me/not me judgment tasks, but there was no

consistent evidence that either group enjoyed a speed advantage in responding to trait terms of interest when response times were standardized. These results suggest that FS are not more schematic than IS on the trait dimensions listed above.

On the other hand, FS did tend to differentiate TAT characters more on the trait dimensions under study than IS, an indication that they projected split object relations onto the neutral TAT stimulus card. FS also showed a greater tendency to respond nonrandomly to the trait terms *good* and *loving* when these were presented subliminally, suggesting a greater degree of priming on these trait dimensions.

The mixed results suggest that although splitting cannot be conceptualized in the schematic terms proposed at the outset of the study, further research into differences in the information processing advantages of FS on the trait dimensions *good*, *bad*, *loving*, and *hateful* is warranted.

TABLE OF CONTENTS

LIST OF TABLES

Table

LIST OF FIGURES

CHAPTER 1

REVIEW OF THE RELEVANT LITERATURE

In this chapter I discuss the Psychoanalytic concept of splitting, distinguishing normal developmental splitting, normal defensive splitting and borderline splitting. I then propose links between splitting and research in social cognition dealing with self-schemata and the complexity of self-representations.

Freud's Splitting of Consciousness

The modern concept of splitting has roots extending at least as far back as Janet, who believed that classical hysteria resulted from a splitting of the contents of consciousness attributable to "an inborn weakness in the capacity for psychical synthesis..." (Freud, 1955a, p.49).

In *Studies of Hysteria*, written between 1893 and 1895, Josef Breuer and Sigmund Freud also recognized a splitting of consciousness in hysteria. They claimed that unlike the everyday "psychical groups" of associated memories, thoughts, and affects accessible from states of normal consciousness, the split-off psychical groups of the hysteric were accessible only within altered states of consciousness--Freud's (1955a) so-called "hypnoid states".

In addition to the remarkable conversion symptoms Anna O. suffered, Breuer and Freud (1955, p.24) described

> ...entirely distinct states of consciousness... which alternated very frequently and without warning and which became more and more differentiated in the course of the illness. In one of these states she recognized her surroundings; she was melancholy and anxious, but relatively normal. In the other state she hallucinated and was 'naughty'--that is to say, she was abusive, used to throw the cushions at people...[and] tore buttons off her bedclothes and linen... There were extremely rapid changes of mood... At moments when her mind was quite clear she would complain of the profound darkness in her head, of not being able to think, of becoming blind and deaf, of having two selves, a real one and an evil one which forced her to behave badly, and so on.[1]

Unlike Janet, for whom split consciousness was essentially congenital, Breuer and Freud viewed it as a secondary manifestation of a more basic defensive process:

> In contradistinction to Janet's view, which seems to me to admit of many and various objections, we have that advocated by J. Breuer in our joint publication. According to Breuer, the "foundation and condition" of hysteria is the occurrence of peculiar dream-like states of consciousness with diminished capacity for association, for which he suggests the name "hypnoid states." The splitting of consciousness is then secondary and acquired; it occurs because the ideas which emerge in hypnoid states are cut off from associative connection with the remaining contents of consciousness (Freud, 1955a, p.49).

---

[1] The reader may detect similarities here with more recent descriptions of multiple personality and dissociative disorders. Unfortunately, the theoretical relationship between splitting, dissociation, and multiple personality formation remains ill-defined a century after Breuer and Freud's collaboration (e.g., Berman, 1981; Armstrong et al., 1990).

In hypnoid states Breuer and Freud (1955, p.12) contended that:

> ...the ideas which emerge... are very intense but are cut off from associative communication with the rest of the content of consciousness. Associations may take place between these hypnoid states, and their ideational content can in this way reach a more or less high degree of psychical organization.

Such splits were not necessarily complete or entirely irreversible:

> The nature of these states and the extent to which they are cut off from the remaining conscious processes must be supposed to vary just as happens in hypnosis, which ranges from a light drowsiness to somnambulism, from complete recollection to total amnesia (Breuer & Freud, 1955, p.12).

Freud soon clarified that splitting of consciousness was not a necessary outcome in hysteria. In *The Defence Neuro-Psychoses* Freud argued that there were three types of hysteria, only two of which resulted in split consciousness. *Hypnoid Hysteria*, already discussed, involved split consciousness secondary to hypnoid states. On the other hand, Freud did not see split consciousness playing any notable role in *pure retention hysterias*, in which the only problem was nonresponse to a trauma, and for which simple abreaction was considered a cure (Freud, 1955a).

As in the case of hypnoid hysteria, Freud's *defence hysteria* did involve splitting of the contents of consciousness, but here it was viewed as the

> consequence of a voluntary act on the part of the patient; that is to say, it is instituted by an effort of will, the motive of which is discoverable. By this I do not of course mean that the patient intends to

> produce a splitting of his consciousness; the patient's aim is a different one, but instead of attaining its end it produces a splitting of consciousness (Freud, 1955a, p.49).

This descriptions of defense hysteria anticipates a very important aspect of the modern concept of splitting. Where there is will, there is also a degree of consciousness. As noted by Masterson (1976, p.57), although

> ...the function of [splitting]... is to keep contradictory primitive affective states [and self and object representations mutually linked with these affective states] separated from each other... *both states remain in consciousness but do not influence each other* [italics added].

This leads to an interesting point. With heavy, continued use of splitting, there is a parallel development of contradictory self-representations, each involving its own distinctive set of cognitive and affective components. When one self-representation is activated, the other is de-activated. Being de-activated, however, does not mean it is completely out of consciousness. To the contrary, a person who is splitting is always at least preconscious of the de-activated self-representation.

This makes splitting a rather shaky defense if the purpose is to avoid awareness of conflict. People who use splitting as a primary defense (e.g., borderlines), are vulnerable to collapse of defensive capacity whenever aspects of the currently de-activated self-representation is brought to full awareness, thus conflicting with a current contradictory self-representation. A typical secondary

defense, at this point, is isolation of affect, in which affect is decoupled from cognition, resulting in denial of the emotional significance of the cognitively acknowledged contradiction (Kernberg, 1976).

Guided by the evolving topographic model, Freud soon turned his energies to the study of repression, a more efficient and adaptive defense than splitting of consciousness. In contrast to splitting of consciousness, repression produced true unconsciousness of unwanted aspects of self or others, while simultaneously permitting partial, symbolic expression of whatever was repressed. Rather than the parallel development of contradictory ego states and self-schemata, repression permitted one to develop a much more coherent sense of self.

Freud's decision to focus on repression was largely strategic; by assuming that most mentally disturbed people had intact egos (and, implicitly, singular conscious selves), Freud was able to push classical drive theory to its limit. By the 1930s, however, many psychoanalytic thinkers realized that some mental disorders--what are now considered character disorders, for example--could not be effectively treated using only strictly traditional psychoanalytic methods. Once the problem was identified as the lack of intact egos in these "unanalyzable patients," psychoanalysis became preoccupied with problems of ego development.

It was in this context that Freud expanded on the original notion of defense hysteria, noting that ego could divide defensively, with opposing memories, thoughts, affects, and fantasies attaching to the part-egos. This position, anticipating modern theories of splitting in the so-called borderline character disorders, reflected Freud's recognition that a defense operation could cause something even more profound than analyzable neurotic symptoms--in essence, a division of the self, or "ego-splitting" (Freud, 1955b, 1955c).

Developmental Splitting

Before discussing defensive splitting in adults, it should be emphasized that most psychoanalytic thinkers believe that normal childhood splitting plays a critical role in the development of cognitive representations of self and others.

By 1921, Melanie Klein believed that at a certain point in development young children split the internal psychic representations of a parent in order to avoid premature contamination of the "good parent" from less savory representations of the same figure (Klein, 1950).

The child's inner world was considered to be as much a source of potential negative associations with parental representations as the real behavior of the parent. As Klein

(1955) later proposed, projective identification, primarily involving the splitting off and projection of the ego's aggressive tendencies, was as central to the so-called paranoid-schizoid stage of development as was splitting, per se.

Klein believed that in permitting the ego to retain a central good object (the "good breast") around which it could ultimately organize benign object relations, normal infantile splitting paradoxically facilitated longer term integration. Normal splitting gave the developing child contrasting self-representations and contrasting object representations with which it could unambiguously associate any of the widely varying phenomenological experiences it encountered. Thus, most of the child's experience could remain object-related.

Klein saw such splits resolving dialectically in a normal developmental course. Normal splitting bought the developing ego time to mature to a point at which its integrative capacity could undo the early splitting and produce unified, ambivalently experienced internalized representations of both self and significant objects--a precondition of the more developmentally advanced depressive position (Klein, 1955).

Unfortunately, the schism between followers of Klein and those of Anna Freud delayed recognition of Klein's important theoretical contributions in the areas of

splitting, projective identification and introjective identification--particularly after Anna Freud failed to mention any of these processes in *The Ego and the Mechanisms of Defence*, originally published in 1936 (Freud, 1948). Loyalty to Anna Freud, combined with abhorrence of some of the more sensational (and often implausible) aspects of Klein's theories, led many of Klein's critics to dismiss her ideas outright (Jones, 1950).

Margaret Mahler, who like Klein worked primarily with young children, also conceptualized early splitting as a normal process, attributing adaptive, as well as defensive functions to it. Her theory of the subphases of separation-individuation stipulated a normal period of splitting in which internal self and object representations are firmly distinguished from one another (Mahler, Pine & Bergman, 1975).

Ego psychology and the British object relations school thus concluded that the most essential component of reality testing--the ability to distinguish internal from external stimuli--was itself a function of normal developmental splitting (Blanck & Blanck, 1974).

Kernberg's elegant litmus test for distinguishing character disorders, psychotic conditions, and neurotic conditions is relevant here. Two key questions must be answered. First, is the patient's reality testing intact? Second, is the patient characteristically dependent upon

"primitive" defenses such as splitting? (Kernberg, 1984, 1988) These questions address the degree to which the patient's normal developmental splitting of childhood was ultimately resolved. Did the patient retain the split between self and object, and did the patient mend the developmental split in self- and object-representations?

Borderline Splitting

In the borderline literature, "splitting" usually refers to the phase-inappropriate and excessive use of splitting defenses clinicians commonly see in adult borderline patients. Kernberg (1975, p.29) provided the classic definition of intrapsychic splitting as the "active process of keeping apart introjections and identifications of opposite quality". Together, contradictory cognitions and affects comprise the "opposite quality" of split introjections and identifications.

Kernberg (1975) claimed that when splitting persists too long in childhood, it short-circuits neutralization of aggressive and libidinal drive, thereby depriving the ego of the sort of moderated and focused energy needed for continuing psychic development. Continually buffeted by untempered drive derivatives (e.g., wishes, fantasies, primitive impulses, etc.) associated with the split self- and object-representations, the ego is unable to establish a

consistent frame of reference from which to mediate between the immediate demands of drive and the ultimate constraints of reality. Kernberg's "ego weakness" refers to the varied impairments of ego function that ensue as developmental failures cascade upon one another.

Discussing adult character disordered psychiatric patients whose cognitive representations of self and objects usually remain split. Kernberg (1975, p.29) noted that:

> The direct clinical manifestation of splitting may be the alternate expression of complementary sides of a conflict in certain character disorders, combined with a bland denial and lack of concern over the contradiction in his behavior and internal experience by the patient... Probably the best known manifestation of splitting is the division of external objects into "all good" ones and "all bad" ones, with the concomitant possibility of complete, abrupt shifts of an object from one extreme compartment to the other; that is, sudden and complete reversals of all feelings and conceptualizations about a particular person. Extreme and repetitive oscillation between contradictory self concepts may also be the result of the mechanism of splitting.

The rapidity and massiveness of the switches from one split perspective to another is what makes borderline patients so memorable to their therapists. Consider one of Harold Searles' patients who told him "You deserve the Congressional Medal of Spit." As Searles (1986, p.501) noted, "The first seven words of that eight-word sentence conveyed heartfelt admiration; but the last one, said with no break at all in the rhythm of her speech, was uttered in unalloyed contempt."

Unlike psychologically healthy adults, who may split selectively, and even adaptively on occasion, severe borderline splitting is essentially uncontrolled and pervasive. Whereas a healthy person may split off relatively small portions of self- and object-representations, the borderline splits self- and object-representations into wildly contrasting halves.

The problem with the borderline, in other words, is structural. Severe borderlines do not, at a basic level, have any conception of the self or objects as whole entities, with both *good* and *bad* aspects. They sometimes experience themselves as a "good self" and sometimes a "bad self". Likewise, they experience significant objects sometimes as "all good", sometimes as "all bad."

As Kernberg (1975) noted, we rarely, if ever, see pure splitting, even with severe borderline patients; instead, this defense combines with infantile forms of idealization and projection, projective identification, denial, omnipotence and devaluation. Acting synergistically with one another, such operations severely undermine cognitive mediation of extreme affect and impulsivity, and make realistic object relations impossible. The combination of splitting and projective identification is especially damaging to borderline patients because it tends to reproduce, in the real relational world, the same chaos and distortion characterizing internal object relations.

In *The Primitive Edge of Experience*, Thomas Ogden spent considerable time outlining the theoretically and experientially amorphous boundary between what might be considered purely intrapsychic splitting processes and their inevitable interpersonal sequelae:

> Splitting defensively renders object-related experience of a given emotional valence (for example, the relationship of a loving self to a loving object) discontinuous from object-related experience of other valences (for example, the relationship of a hating self to a hating object). Each time a good object is disappointing, it is no longer experienced as a good object- not even as a *disappointing* good object- but as the discovery of a bad object in what had been masquerading as a good one. Instead of ambivalence there is the experience of unmasking the truth. This results in a continual rewriting of history such that the present experience of the object is projected backward and forward in time creating an eternal present that has only a superficial resemblance to time as experienced in a... [more conventional] mode... Rewriting of history leads to a brittleness and instability of object relations that are in continual states of reversal (Ogden, 1989, p.19).

It takes little imagination to envision the destructiveness of such a phenomenological position in the interpersonal realm. In a study of severely ill borderline patients Greene, Rosenkrantz and Muth (1986, p.256) noted that:

> By means of splitting operations, good and bad, idealized and devalued, internal objects are forcefully dissociated from each other; the borderline's internalized social schemata are thought to consist of rigidly dichotomized and polarized categories- pleasing and need-satisfying versus threatening and need-thwarting - of interpersonal experience. To reinforce these emotionally crude distinctions, loved and hated internalized objects are continuously and massively projected onto and into different segments of the current social field... the borderline basically is not

> invested in the painstaking work of discovering the reality of actual social objects; the paramount need is to find external figures capable of being transformed magically and immediately, via projective endowments, into the good and bad, protective and haunting objects of the borderline's internal world...

Viewing objects alternatively as all good or all bad, and lacking more subtle cognitive schemata with which to realistically appreciate actual objects, borderline patients essentially try to make the world conform to their sense of reality so as to maintain some sense of order, predictability and meaning for themselves. Through such unconscious mechanisms as projective identification, borderline patients often manage to induce in real objects the very affects, moods, impulses, fantasies, and other cognitions unconsciously "disowned" through splitting (Cashdan, 1988; Ogden, 1982, 1986, 1989; Scharff, 1992). A retraumatizing positive feedback loop results as the borderline's unrealistic split self- and object-schemata, and the behavior they give rise to, unwittingly induce reactions in the interpersonal field which reinforce the borderline's distorted view of that field.

One well known example of splitting acting in conjunction with projective identification occurs in the phenomenon referred to as "staff splitting" of inpatient hospital staff by borderline patients. In "staff splitting" a number of unit workers are unconsciously induced by the borderline patient, through projective identification, to act out certain split-off aspects of the borderline's ego.

Some staff members, for example, may find themselves experiencing the split-off rage of the patient, toward whom they may assume a rather punitive position. Other staff members, at the same time, may find themselves experiencing an overwhelming need to protect and defend the patient. Unless the situation is correctly diagnosed as a manifestation of the patient's projective identifications, the situation may easily escalate to the point at which staff spend increasing amounts of time passionately disagreeing about which treatment approach to follow. They may even sabotage each other's work (Gabbard, 1989).

Detecting Splitting

Students of splitting have thusfar taken four general approaches to the problem of detecting splitting- which like many other intrapsychic processes can only be inferred on the basis of observable behavior.

The first method, the clinical method, is the oldest of the four approaches and the most crucial in psychotherapeutic work. The clinical approach is exemplified by the clinician who infers the operation of splitting after observing rapid and dramatic shifts in a patient's expressed attitude toward significant objects or the self (for example, recall the quote from Searles' patient).

An important advantage of the clinical method of detecting splitting is that it allows the possibility of a clinician becoming aware of splitting almost immediately, thereby making timely intervention conceivable. Assuming that the clinician has become well attuned to the various nuances of the patient's dynamics and communications, the clinical method may also achieve high reliability and validity in particular clinical situations.

Unfortunately, reliability cannot be assumed in every case. There are at least three important problems. First, there is the issue of calibration. What degree of change in one's expressed attitudes toward self or others is required before it is reasonable to infer a splitting operation? What one clinician hears as a disjunctive shift in emotional tone and attitude (suggestive of splitting) might be perceived by another clinician as less disjunctive and more suggestive of ambivalence.

Second, verbally communicated manifestations of splitting are probably influenced to some extent by such things as the splitter's educational level and the degree to which the splitter has been socialized to speak strong thoughts directly, or to withhold them.

Finally, countertransference is notoriously powerful in work with patients who split frequently. When splitting is combined with projective identification and idealization or devaluation of the therapist, the clinician's own

narcissistic needs and vulnerabilities may be directly tapped. Sometimes, in such cases, splitting goes undiagnosed while the clinician resonates uninsightfully with the patient's distorted views of him or her.

In summary, the clinical method of detecting splitting, dependent as it is on the training and experience of the clinician, the quality of the relationship between the patient and clinician, the vicissitudes of the clinician's attention to the possibility of splitting, and many other variable and essentially unquantifiable factors, must be considered insufficiently rigorous for serious scientific purposes.

The second general method for measuring splitting involves the use of self-administered scales, like that developed by Gerson (1984) to identify people with a characterlogical propensity for splitting. Gerson's 14 item scale (Table 1), based on her survey of splitting literature in ego psychology and self psychology, included items designed to assess such things as separation of good and bad images of self and other, anger, idealization, grandiosity, and Kernberg's (1975) "identity diffusion", all theoretically linked to splitting.

An interesting and important aspect of Gerson's scale (or any self-administered defense scale) is that it requires subjects to be aware of manifestations of what is largely an unconscious defensive operation. Citing Kernberg (1975) and

Kohut (1971), Gerson (1984, p.158) addressed the issue directly, stating that "self-report is an appropriate mode of assessing the defense of splitting because the manifestations of the defense, such as shifting object evaluations, are part of awareness, although the purpose served by the shifts is not."

Gerson (1984) carried out preliminary psychometric assessment of the scale and found scale scores positively associated with a scale measure of narcissistic personality and negatively associated with a self-esteem measure. Glassman (1986) performed additional psychometric work, concluding that seven items were superfluous. Items retained in the revised scale are flagged in Table 1.

The Gerson scale is not intended to detect particular instances of splitting, but only to measure people's proclivity for splitting. If found reliable and valid, such a scale has importance in research as a tool with which to discriminate frequent and infrequent splitters. This is how it was used in the current study, and additional evidence was produced supporting its validity in Study 2.

A third general approach to measuring splitting is exemplified by the Lerner and Lerner (1982) Rorschach scale. Believing that the natural ambiguity of the blots would elicit characteristic defensive reactions in susceptible subjects, the Lerners thought that splitting would manifest in several ways. For example, in a Rorschach response

Table 1. Gerson's (1984) splitting scale.

(Not at all true) 1 2 3 4 5 6 7 (Very True)
1.) I hate to hear someone close to me being criticized.
2.) When I'm with someone really terrific, I feel dumb.[a]
3.) When I'm angry, everyone around me seems rotten.[a]
4.) My friends don't know how much I'd like to be admired by people.
5.) It's hard for me to get angry at people I like.
6.) It's very painful when someone disappoints me.[a]
7.) I have absolutely no sympathy for people who abuse their children.
8.) Sometimes I feel I could do anything in the world.
9.) There are times my wife (husband)/girlfriend (boyfriend) seems as strong as iron, and at other times as helpless as a baby.
10.) I often feel that I can't put the different parts of my personality together, so that there is one *me*.[a]
11.) I sometimes feel my love is dangerous.[a]
12.) When I'm in a new situation, there's often one person I really dislike.[a]
13.) It's hard for me to become sexually excited when I'm depressed.[a]
14.) Some people have too much power over me.[a]

[a]Item retained in Glassman's (1986) revised scale.

"an implicitly idealized figure" could be "tarnished or spoiled by the addition of one or more features or an implicitly devalued figure" could be "enhanced by the addition of one or more features..." Alternatively, a response might involve "two clearly distinguished figures" with "each figure... described in a way opposite to the other: 'two figures, a man and a woman. He is mean and shouting at her. Being rather angelic, she's standing there and taking it.'" (Lerner & Lerner, 1982, pp.85-86).

Table 1. Gerson's (1984) splitting scale.

(Not at all true) 1 2 3 4 5 6 7 (Very True)
1.) I hate to hear someone close to me being criticized.
2.) When I'm with someone really terrific, I feel dumb.[a]
3.) When I'm angry, everyone around me seems rotten.[a]
4.) My friends don't know how much I'd like to be admired by people.
5.) It's hard for me to get angry at people I like.
6.) It's very painful when someone disappoints me.[a]
7.) I have absolutely no sympathy for people who abuse their children.
8.) Sometimes I feel I could do anything in the world.
9.) There are times my wife (husband)/girlfriend (boyfriend) seems as strong as iron, and at other times as helpless as a baby.
10.) I often feel that I can't put the different parts of my personality together, so that there is one *me*.[a]
11.) I sometimes feel my love is dangerous.[a]
12.) When I'm in a new situation, there's often one person I really dislike.[a]
13.) It's hard for me to become sexually excited when I'm depressed.[a]
14.) Some people have too much power over me.[a]

[a]Item retained in Glassman's (1986) revised scale.

"an implicitly idealized figure" could be "tarnished or spoiled by the addition of one or more features or an implicitly devalued figure" could be "enhanced by the addition of one or more features..." Alternatively, a response might involve "two clearly distinguished figures" with "each figure... described in a way opposite to the other: 'two figures, a man and a woman. He is mean and shouting at her. Being rather angelic, she's standing there and taking it.'" (Lerner & Lerner, 1982, pp.85-86).

administering a Rorschach correctly and reliably requires considerable training and practice. Second, the administration itself can be fairly time consuming, and is usually done with one subject at a time. Third, reliable scoring of a Rorschach protocol using the defense scales requires additional training and practice. These factors alone make the use of the Rorschach impractical for most research purposes.

There is also one important theoretical problem in using any projective measure. In the words of Rapaport, Gill, & Schafer (1975, p.224), the projective hypothesis, upon which all projective interpretation is ultimately based, is essentially the idea that "manifestations of the human being's behavior, from the least to the most significant ones, are revealing of his personality." The validity of this hypothesis is far from universally accepted. Thus, studies relying too heavily upon Rorschach measures, or other projective test data, suffer a loss of credibility among many scientifically oriented psychologists.

Greene, Rosenkrantz and Muth (1985) exemplify a fourth means of measuring splitting, focusing more on interpersonal than intrapsychic splitting. Here, evidence of interpersonal splitting is used to infer intrapsychic splitting. The researchers asked members of a borderline therapy group to judge themselves on the dimension of goodness/badness in

relation to other group members, and then asked them to evaluate the group coleaders.

The key finding was that the more a patient rated him/herself as bad in relation to the rest of the group the greater was the tendency to rate the cotherapists differentially on a number of personal dimensions. This suggests that the more extreme the view of the self on the good/bad dimension, *vis a vis* other people with whom one is interacting, the more likely it is that one will use also evaluate others more extremely on salient dimensions.

Interestingly, Morrison, Greene, & Tischler (1985) demonstrated that mental health professionals in Tavistock group training show the same tendency to evaluate group leaders differentially. The Tavistock results can be interpreted as reflecting interpersonal and intrapsychic splitting, only if one makes the additional assumption that the (presumably) mentally healthy mental health workers split defensively, in response to the extraordinary stress of the Tavistock process (Bion,1959), rather than for deeper structural reasons. That is, they regressed to splitting under the pressure of group processes, rather than splitting being a common feature of their psychic life.

This assumption is not implausible, but does suggest the importance of distinguishing evidence of splitting in discrete circumstances and evidence of splitting in a broader context due to characterlogical issues.

Self-Schemata

Although not typically interested in splitting, per se, social psychologists have long studied the development, maintenance, and implications of strongly held views of the self and others. In her landmark study of self-schemata Hazel Markus (1977, p.63) noted that:

> The quantity and variety of social stimulation available at any time is vastly greater than a person can process or even attend to. Therefore, individuals are necessarily selective in what they notice, learn, remember, or infer in any situation. These selective tendencies, of course, are not random but depend on some internal cognitive structures which allow the individual to process the incoming information with some degree of efficiency. Recently, these structures for encoding and representing information have been called... *schemata*...The influence of cognitive structures on the selection and organization of information is probably most apparent when we process information about ourselves. A substantial amount of information processed by an individual (some might even argue a majority of information) is information about the self, and a variety of cognitive structures are necessarily involved in processing this information... It is proposed here that attempts to organize, summarize, or explain one's own behavior in a particular domain will result in the formation of cognitive structures about the self or what might be called self-schemata. *Self schemata are cognitive generalizations about the self, derived from past experience, that organize and guide the processing of self-related information contained in the individual's social experiences.*

From this view self-schemata play a fundamental role in shaping one's constantly evolving sense of reality as it relates to the self, by influencing what one attends to, what one encodes in memory, and the inferences drawn from environmental events (Markus & Nurius, 1986).

To test her ideas, Markus (1977) asked subjects to rate how dependent they were. She defined subjects who rated themselves extremely dependent, or extremely independent as schematic on the dependent/independent dimension, and others who rated themselves more moderately as aschematic.

She then demonstrated, among other things, that being schematic (at either extreme) significantly increased the efficiency with which judgments about the self on the schematic dimension were made. She concluded that people can make much more confident judgments about the self, more rapidly, when queried about schematic dimensions, rather than dimensions which are, for them, aschematic.

Markus' conclusions about the information processing advantages of self-schema use have generally been supported by other research. Most notable, perhaps, is the work of Kuiper (1981), who had subjects rate how much each of a series of adjectives did or did not describe the self, on a 9-point scale[2]. Later, subjects were asked to make a dichotomous judgment whether each adjective was like or unlike them, and response time was recorded. The essential finding was that response latencies were considerably shorter for adjectives rated extremely on the 9-point scale, that is, as very much like the self or unlike the self. The same pattern, which Kuiper graphically referred to as the

---

[2] Kuiper treats self as a prototype, rather than a schema, a technical difference of importance for some purposes, but not here.

"inverted-U RT effect", held when subjects were asked to judge another person according to the same adjectives.

Two important conclusions follow. First, people find it easiest to recognize near-perfect (and thus obvious) matches between the self-concept and adjectives used to describe it, or extremely bad (and likewise obvious) mismatches, but require more information processing effort in judging cases that fall between the extremes. Second, people most efficiently judge others on dimensions most salient to their own sense of self.

What is important is self-concept, not the "objective self". Self-schemata are not necessarily accurate representations of the self; they do not need to be in order to have significant information processing effects. That is, people process information about the self based on whatever self-representations are dominant, and these may or may not bear a lot of resemblance to more objective representations of self.

This simplifies the problem of determining how schematic people are on given dimensions. We do not need to measure how dependent a person really is to judge his or her schematicism on the independence/dependence dimension. Rather, as noted by Fiske and Taylor (1991, pp.183-184), we only need to ask him or her:

> There are several criteria for deciding if someone has a schema (*is schematic*) or has no schema (*is aschematic*) on a particular dimension of his or her

> self concept. People are self-schematic on dimensions that are important to them, on which they think of themselves as extreme, and on which they are certain the opposite does not hold... Thus, if independence is important to you, and if you think of yourself as extremely independent and as not at all dependent, that implies that you have accumulated considerable knowledge about yourself on that dimension. For example, you should be certain that you would never ask for help setting up your stereo, even at the potential cost of damage to it or harm to yourself... In contrast to schematics, aschematics are not invested in, or concerned about a particular attribute. They rate the trait as low to moderately self-descriptive, and perceive it as low to moderate in importance. Of course, people who are aschematic on one trait may well be schematic on others; everyone has some dimensions of self that are idiosyncratically salient.

The second experiment of the present research uses the three criteria mentioned by Fiske and Taylor for determining the degree to which subjects are schematic on a variety of dimensions. The first two criteria, the degree to which subjects rate themselves extreme on a dimension and the degree to which they find that dimension personally significant, logically apply equally well to frequent and infrequent splitters.

However, even though the third criterion (disavowal of the opposite trait) seems eminently appropriate in judging garden variety schematicism, its applicability in evaluating schematicism of split dimensions requires some discussion.

Although the classic sort of self-schema described by Fiske and Taylor is unipolar, self-concept in borderline splitting is bipolar. That is, in the case of a classic self-schema a person endorses one extreme end of a self-dimension, and never endorses the opposite extreme. In

characterlogical splitting, on the other hand, both extremes are endorsed, though not at the same time.

It has long been noted by clinicians, however, that borderline patients are quite capable of acknowledging aspects of self opposed to the currently activated side of a split, when confronted, but give the opposing aspects no emotional, and thus no practical weight (Kernberg, 1975; Masterson, 1976).

Because splitting does not make aspects of self opposed to the currently activated side of the split entirely unconscious (Masterson, 1976), the decision to apply the third Fiske and Taylor (1991) criterion hinged on the following logic: admission of the inactivated side of a split, which generally occurs in response to confrontation in clinical situations, represents at least a partial failure of the splitting defense. The purest cases of splitting should involve denial, to a considerable degree, of the inactivated side of the split.

Thus, by including the third criterion in my measurement of schematicism on dimensions reputed to be important among frequent splitters (borderlines being the most frequent splitters), I hoped to detect the most robust instances of splitting. As described in chapter 2 and chapter 3, the questions subjects answered to yield my overall measure of schematicism were posed prior to other experimental manipulations or challenges. I used this

procedure to maximize the chance that measured schematicism on global dimensions signified by such words as *good*, *bad*, *loving*, and *hateful* would reflect characterlogical splitting, rather than atypical splitting resulting from unusual stress caused by the experimental procedures themselves.

It should be re-emphasized, however, that denial that the opposite ever holds is not necessarily a stable feature of schematicism in borderline splitting. This means that the concept of schematicism, as formally defined by Fiske and Taylor (1991), may only apply within certain parameters in the case of borderline subjects.

Complexity of Self-Representations

The Fiske and Taylor (1991) quotation implies that self-schemata apply to somewhat limited dimensions of self. As a corollary, one can infer that a person can be schematic on several relatively independent dimensions. For example, it is possible to be a schematic sports fan, a schematic Democrat, and a schematic gourmet, with each schema remaining fairly independent of the others.

In order to conceptualize borderline splitting in terms of self-schemata, however, one must imagine that the dichotomous self-schemata are a great deal more global. That is, rather than having the normal complement of reasonably

independent self-schemata associated with circumscribed dimensions of self, the borderline is conceptualized as having two alternatively activated, self-schemata, both of which are relatively all-encompassing.

By all-encompassing I mean that the good and bad self-schemata of the borderline subsume most other aspects of self. This implies that they are also salient in most life situations. Although a great many life events have no relevance to a Democrat self-schema, or a sports fan self-schema, almost any life event has potential relevance to a hypothetical good self-schema or bad self-schema.

Linville (1985, p.95) described people as having complex representations of self or others when they organize self-knowledge "in terms of a greater number of aspects that are relatively independent of one another." She then demonstrated that complexity of self- or other-representation is associated with less extreme evaluations of self and others, and less affective extremity. In her view more complex representations serve to buffer evaluations and affect by providing a framework for integrating discrepant bits of information about self or others (Linville, 1982, 1985, 1987).

The borderline good self- and bad self-schemata must be considered simple representations in Linville's sense, because they involve two global evaluations, with but one active at any given time. To the extent they subsume any

less global aspects of self, the more circumscribed aspects of self-representation are, by definition, highly correlated under the good or bad rubrics. According to Linville's theory, the simple self-representations we see with borderlines should result in extreme evaluations of the self as well as extreme affect associated with such appraisal.

Affect and Self-Schemata

The literature dealing with depressive self-schemata (e.g., Beck, 1979; Kuiper et al., 1985) exemplifies a broader effort to understand emotions from a cognitive perspective. From this perspective the affect elicited by an event depends as much upon personal "meaning-making" as the event's objective aspects.

Beck's (1979, p.84) "cognitive triad" of depression, which he implicates in the development of most depressive disorders, includes "a negative conception of the self, a negative interpretation of life experiences, and a nihilistic view of the future." The cognitive triad is pertinent to the present study because the "negative conception of the self" is essentially a self-schema which is associated quite closely, in Beck's view, with a particular affective experience.

Fiske (1982) empirically demonstrated the close linkage of affect and certain schemata by showing that when a schema

is activated, the affect associated with it is also triggered. One must be careful not to assume that the opposite necessarily holds, however. In the normal situation a particular affect could be associated with any number of schemata, and one couldn't reasonably expect affect to trigger a particular schema just because the schema triggered the affect. What works in one direction may not work in reverse.

The situation in borderline splitting is hardly normal, however. Here the world and the self are split into two overarching categories--good and bad. Given this, it is not unreasonable to expect euthymic mood (i.e., positive or "good" mood) to trigger the good schema, and dysthymic mood the bad schema. If, as Linville (1987) noted, the less complex a person's sense of self, the larger the proportion of self-aspects likely to be affected by an emotionally salient event, a borderline's simple sense of self should be massively affected by any induced mood change.

This fits current wisdom regarding borderline patients. After noting that borderline individuals are subject to emotional dysregulation, characterized by a low threshold for emotional reaction, intense emotional reaction once the threshold is passed, and a slower than normal return to emotional baseline, Linehan (1993) emphasized that strong emotion narrows attention to mood-congruent information, so that cognitions tend to fall into line with strong affect.

This suggests that once strong affect is triggered in a borderline, splitting is liable to occur; the good self or the bad self is elicited or reinforced. Whether the split will necessarily be mood congruent, or defensively mood incongruent (with the mood congruent side of the split perhaps projected onto an object) is not so certain, however.

Splitting and Self-Schemata

The present research grew out of the idea that splitting can be understood as a process in which opposing self-schemata are alternatively activated. A related premise holds that any splitting (even if it is not of the alternating borderline variety) involves activation of a self-schema.

In social cognition an active self-schema is believed to manifest in both the willful (i.e., conscious) and automatic (i.e., unconscious) behavior of a person. In the willful realm, as noted above, a person acting schematically will overtly identify him or herself as extreme on schematic dimensions, will assert the personal importance of this identification, and will also assert that an opposite identification does not hold true.

Two methods were employed in the two studies described below to tap the conscious side of self-schematicism. First, subjects were asked to locate themselves on scales representing a trait of interest. This provides a measure of a subject's extremity of self-concept on that trait dimension. Second, subjects were asked to judge, as rapidly as possible, whether trait adjectives presented on a computer monitor were best characterized as *me* or *not me*.

A person should also show automatic (unconscious) improvement in ability to process schema-related information in comparison with aschematic information. In the second experiment of the present study subjects were asked to "guess" whether subliminally presented trait adjectives did or did not describe them. Because they remained consciously unaware of the words presented, significant variance from a random response pattern on particular terms provided one measure of automatic improvement in information processing, in this case manifesting as an improved ability to "guess" consistently, presumably in congruence with an underlying schema which had primed them to respond nonrandomly.

Lowered response latency in me /not me judgments about adjectives presented on either subliminal tasks or superliminal tasks was treated as another potential indicator of improved information processing.

Summary

In this chapter I outlined the historical development of the concept of splitting in the psychoanalytic tradition, and distinguished between normal developmental splitting and borderline splitting. I then described several general methods which have been used to detect splitting for clinical or research purposes.

Next I discussed the concept of self-schemata from the social cognitive school, with emphasis on improvements in information processing associated with the use of self-schemata. Finally, I speculated that splitting can be usefully described in terms of extreme, global, and alternating self-schemata.

CHAPTER 2

STUDY ONE

In the fall of 1993 I conducted pilot research intended to accomplish two primary goals: 1) to explore the use of mood inductions as a means of inducing splitting, and 2) to consider the usefulness of inconsistencies in self-evaluation, under contrasting moods, as a marker of splitting.

In this chapter I describe the experiment in detail, present and analyze the results, and discuss changes in experimental design prompted by the mixed results.

General Hypotheses

Assuming that splitting could be provoked in susceptible subjects using the mood induction procedure described by Parrott (1991), described in detail below, I expected the splitting to manifest as inconsistency in judgments about the self-descriptiveness of adjectives after subjects experienced contrasting mood inductions. The general logic is simple: if splitting involves a switch from one self-schema to another then judgments about the self-descriptiveness of trait terms central to those self-schemata should vary as the schemata alternate.

Because frequent splitters (as determined by the Gerson Splitting Scale) were expected by definition to split more than infrequent splitters during the experiment, as a result of contrasting mood inductions, I also predicted that frequent splitters would show more inconsistency than those identified as infrequent splitters.

Another major hypothesis was that subjects identified as frequent splitters would rate the self-descriptiveness of trait adjectives in a more extreme fashion than subjects identified by their Gerson scale scores as infrequent splitters. I based this hypothesis on the supposition that people who split characterologically (and thus with greatest frequency) have simple bipolar self-representations and consequently evaluate self and others in a more extreme way.

Note that extremity of self-rating is not necessarily indicative of schematicism on a trait, because the rating may reflect the view that one is entirely unlike a trait. Henceforth I will use the term *schematic* to refer to situations in which a trait is strongly endorsed as self-descriptive, the term *antischematic* to refer to situations in which a trait is strongly denied to be self-descriptive, and the term *aschematic* in situations in which a trait is neither strongly endorsed as self-descriptive nor denied as self-descriptive.

Finally, I hypothesized that frequent splitters would tend to respond more rapidly than infrequent splitters, when

asked to judge the self-descriptiveness of the terms good and bad, on the theory that these terms tap the extreme, contradictory self-schemata which should be more characteristic of frequent than infrequent splitters.

Methods

In this section I describe the experimental procedures in detail.

Sampling

Study 1 and Study 2 were labor intensive, requiring my assistants and I to spend about an hour with each subject. Thus, I found it necessary to limit the experimental sample to the smallest size offering reasonable statistical power. Because frequent splitting is an uncommon characteristic in the general population, however, a small random sample would have yielded few, if any, frequent splitters. Thus, I employed disproportional stratified sampling to yield two subsamples with extreme scores on the splitting scale.

Thanks to a prescreening process psychology majors are required to undergo, we were able to administer the revised Gerson scale (Glassman, 1986) to more than 1,000 undergraduates.

Prescreening data were written to an SPSS file on the University's mainframe computer. I sorted and stratified these student records by Gerson scale score, and combined the highest strata with the lowest, the next highest with the next lowest, and so on. In this way I generated new strata of scores ranging from most extreme (high and low) to most moderate. I retained only the most extreme strata for further use, randomizing records within strata.

I produced a list of student names and phone numbers, with no indication of Gerson scale score (other than degree of extremity). My assistants and I called students in the order they were listed, and asked them to participate in an experiment[3]. The only substantive information we provided about the experiment was that it involved putting oneself in the mood suggested by certain pieces of music. More than fifty students agreed to participate but only 39 ultimately participated (19 frequent splitters and 20 infrequent splitters). The rest failed to show up as scheduled.

Study Procedures

A researcher met with individual subjects for approximately an hour in a small quiet, windowless

---

[3] A separate list of all potential subjects, sorted by phone number, was maintained, checked regularly throughout the sampling process to insure that roommates were not sampled. This was a precaution against subjects hearing about the experimental design before coming in.

experimental room in Tobin Hall, at the University of Massachusetts. The room was furnished with two chairs and a desk. Items on the desk included a computer monitor and keyboard hooked up to an IBM type PC, a cassette tape deck, and a pair of stereo headphones capable of playing prerecorded cassette tapes with a high degree of fidelity.

We asked subjects to read and sign an informed consent form which explained, among other things, that there was some small risk they might experience unforeseen effects as a result of the mood induction procedure. All subjects signed the release. We then oriented them to the equipment on the desk.

In particular, we directed their attention to a cover which fit over the keyboard in such a way that only two groups of function keys and another set of keys labelled with the digits from 1 to 9 were accessible. The first group of function keys, comprised of four horizontally aligned keys, was labelled *not me* on the keyboard cover. A second group of four horizontally aligned function keys, located five inches to the right of the first group, was labelled *me*.

Thus, we explained, subjects would be able to respond *me* by pressing any one of the four keys so labelled, or *not me* by pressing any one of the other four accessible keys.

We started the computer program and it prompted subjects to indicate the degree to which various adjectives

did or did not describe them, using the digit keys from 1 to 9 (with 1 indicating complete rejection of the adjective as self-descriptive and 9 representing complete endorsement of the adjective). Gough's (1952) Adjective Check provided 14 of the 36 adjectives used in this study, with others suggested by friends and colleagues who helped test the computer program (Table 2).

Table 2. Adjectives used in Study one.

Ambitious[a]	Happy	Sinful[a]
Assertive[a]	Humble	Smart
Athletic	Impulsive[a]	Stable[a]
Attractive[a]	Insecure	Stupid
Bad	Intellectual	Thin
Believable	Kind[a]	Timid[a]
Careful	Liberal	Ugly
Clumsy	Moral	Unworthy
Confident[a]	Nervous[a]	Wanted
Emotional[a]	Rational[a]	Weak[a]
Generous[a]	Right	Well-liked
Good	Sad	Wrong

[a]Derived from Gough (1952).

It should be emphasized that this was a rather primitive measure of schematicism, incorporating only the first of the three Fiske and Taylor (1991) criteria for determining whether a person is schematic on a given dimension.

During a pause in the computer program, we asked subjects to listen carefully to a musical selection, and to put themselves in the mood suggested by the music. All subjects heard Delibes' *Mazurka* from *Coppélia*, at this time

(Delibes et al., 1989). This music, lasting approximately four and a half minutes, was intended to induce happy mood.

Musical mood induction is one of several approaches taken in recent research to alter the mood of experimental subjects (Clark, 1983; Kenneally, 1988). Parrott (1991) specifically validated Delibes' piece for happy mood induction and Prokofiev's stark *Russia under the Yoke of the Monguls*, from the soundtrack of Eisenstein's *Alexander Nevsky*, for sad mood induction (Prokofiev et al., 1987).

The computer program was restarted and subjects were then instructed to judge, *as rapidly as possible*, whether the adjectives they would soon see on the screen did or did not describe them, and to respond either *me* or *not me* by pressing an appropriate key. The computer displayed the adjectives against a black screen in white letters 1.3 cm high, from a distance of approximately 61 cm. Adjectives were presented in unique random order for each subject, one at a time, for a maximum of 6 seconds. After 6 seconds the computer scored the response *missing* and cycled to the next adjective. The computer automatically recorded response latency and direction of successful responses.

Next, on the basis of random assignment unknown to the experimenter until that moment, subjects listened to a repeat of the Delibes piece or, if the subject was assigned to the experimental, contrasting mood induction condition, to Prokofiev's *Russia, under the Yoke of the Monguls*.

The latter piece was not played as half speed, as Parrott (1991) presented it to his subjects, because this would have extended its normal three and three quarters minutes length to seven and a half minutes, exceeding the time which could be allotted for mood induction. Even at regular speed the music is grimly evocative, as aptly described by music reviewer Steven Ledbetter (1987, p.3):

> The film opens on a scene of desolate empty steppes littered with signs of past battle. Prokofiev's musical equivalent of this desolation is a keening melody played in unison, four octaves apart, with noting but emptiness in between. The oboes' lamenting tune suggest poignant loss, while the rapid turn figure in the muted violas and violins is an image of the feather-grass blowing on the hillside--the only thing moving.

Following the second mood induction, subjects repeated the adjective task, with adjectives presented in a new random order. Again, the computer recorded direction of response and response latency. After this we administered the happy mood induction to subjects who had received the sad mood induction. We debriefed all subjects fully, paying special attention to any emotional changes the subjects reported (Appendices B and C). We then dismissed subjects.

Specific Hypotheses

Specific hypotheses tested in this study are listed below, their rationale having already been discussed.

Hypothesis 1

Hypothesis 1 states that the mean consistency rate of me/not me judgments will be higher for subjects with high Gerson scale scores than subjects with low Gerson scale scores.

Hypothesis 2

Hypothesis 2 states that the mean consistency rate of me/not me judgments will be higher for subjects in the experimental condition than subjects in the control condition.

Hypotheses 3a and 3b

Hypothesis 3a states that subjects with high Gerson scale scores will deviate further than subjects with low Gerson scores from the midpoint value on the 9-point scale on which they are asked to rate the self-descriptiveness of the trait term *good*.

Hypothesis 3b states that subjects with high Gerson scale scores will deviate further than subjects with low Gerson scores from the midpoint value on the 9-point scale on which they are asked to rate the self-descriptiveness of the trait term *bad*.

Hypotheses 4a and 4b

Hypothesis 4a predicts subjects with high Gerson scale scores will respond quicker in making judgments about the trait adjective *good* than subjects with low Gerson scores.

Hypothesis 4b predicts subjects with high Gerson scale scores will respond quicker in making judgments about the trait adjective *good* than subjects with low Gerson scores.

Results

This section presents experimental results pertinent to the hypotheses.

Hypotheses 1 and 2

Hypothesis 1 states that the mean consistency rate of me/not me judgments will be higher for subjects with high Gerson scale scores than subjects with low scale scores.

The 39 subjects produced 1342 usable pairs of matched adjective responses, one response to the adjective following the first mood induction, and one following the second mood induction (an additional 62 pairs of responses were spoiled by nonresponse to the adjective on one or both trials). We scored these pairs as *consistent* if both responses were identical, or *inconsistent* if there was a change from Trial 1 to Trial 2.

Figure 1 compares the mean consistency rate of me/not me judgments by subjects with low Gerson scores and subjects with high Gerson scores. This difference of means is in the direction predicted by hypothesis 1. The differences are not statistically significant, however ($t(22.1) = -1.055$, $p=.15$), and the null hypothesis for Hypothesis 1 cannot be rejected.

In examining the data, I discovered that some response times for me/not me judgments were so short (e.g., one was only 2 ms!) that they called into question the legitimacy of the consistency score for that item. Removing items from analysis if subjects responded in less than 100 ms yielded a mean consistency rate of .93 for infrequent splitters and .86 for frequent splitters, a difference closely approaching significance ($t(25.3) = -1.565$, $p=.065$).

Hypothesis 2 states that the mean consistency rate of me/not me judgments will be higher for subjects in the experimental condition than subjects in the control condition.

Figure 2 compares the mean consistency rate of me/not me judgments of subjects in the control condition versus those in the experimental condition. The difference of means is in the direction predicted in hypothesis 2, but the results are not statistically significant ($t(23.1) = .831$, $p=.20$). Taking out of analysis responses of less than 100 ms did not appreciably alter observed mean differences or their statistical significance.

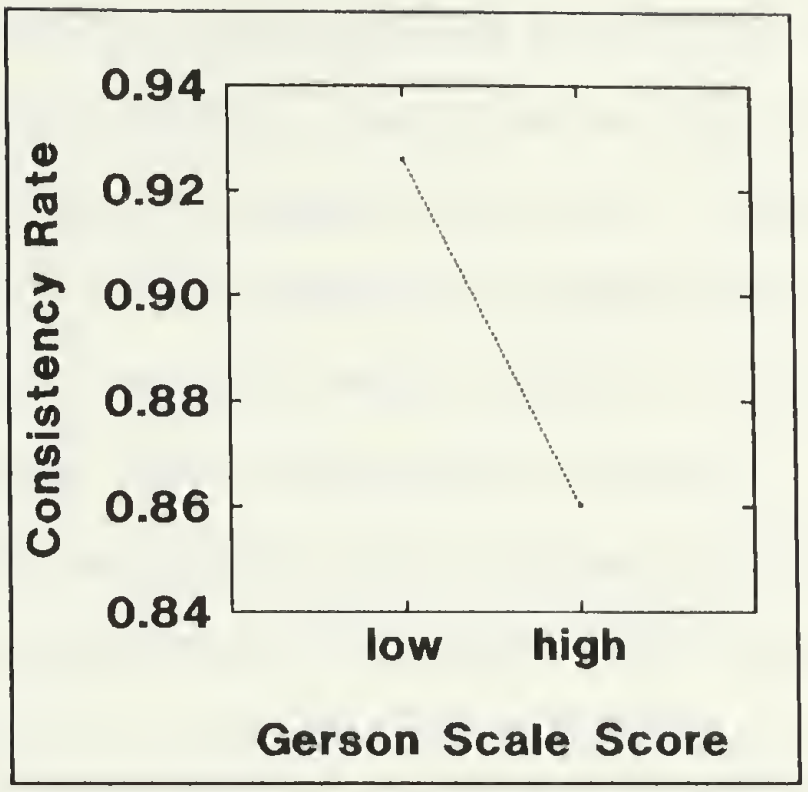

Figure 1. Mean consistency rate of me/not me trait judgments by Gerson scale score groups.

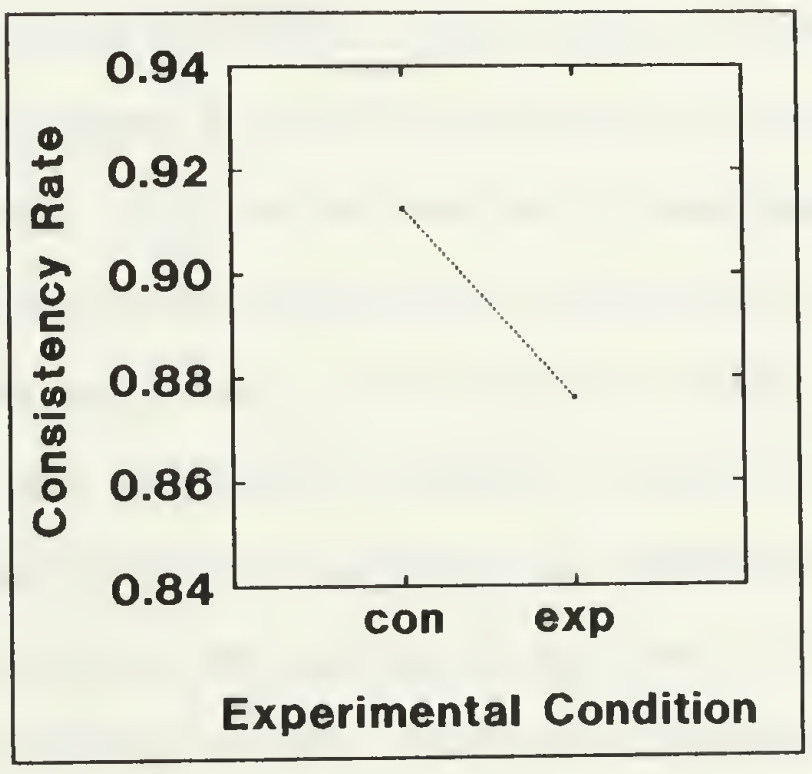

Figure 2. Mean consistency rate of me/not me trait judgments by control (Con) and experimental (Exp) conditions.

Figure 3 depicts the combined effects of propensity for splitting and experimental condition on the consistency rate of me/not me judgments. Despite the fact that the trend appears precisely that predicted by Hypothesis 1 and Hypothesis 2, none of four means differ to a statistically significant degree when contrasted in pairs.

It may seem surprising that having observed an almost statistically significant difference between the consistency rates of frequent and infrequent splitters, we would fail to approach significance when contrasting the most different mean consistency rates of subgroups of the frequent and infrequent splitters. The explanation is simple: as the size of contrasted groups declines so does the power of the *t*-test as an inferential tool. An observed difference of means which approaches significance when N=39 is liable to be far from significant when N is only half that large.

In retrospect, the experimental arrangement, and the nature of the data, did not permit sufficiently powerful statistical testing; one would normally expect mean differences of the magnitude observed above to achieve statistical significance. Although 39 subjects would be sufficient for many research designs, the large variance of many variables in the present data set, and the allotment of subjects to four distinct groups, combine to reduce power of the statistical procedures available here.

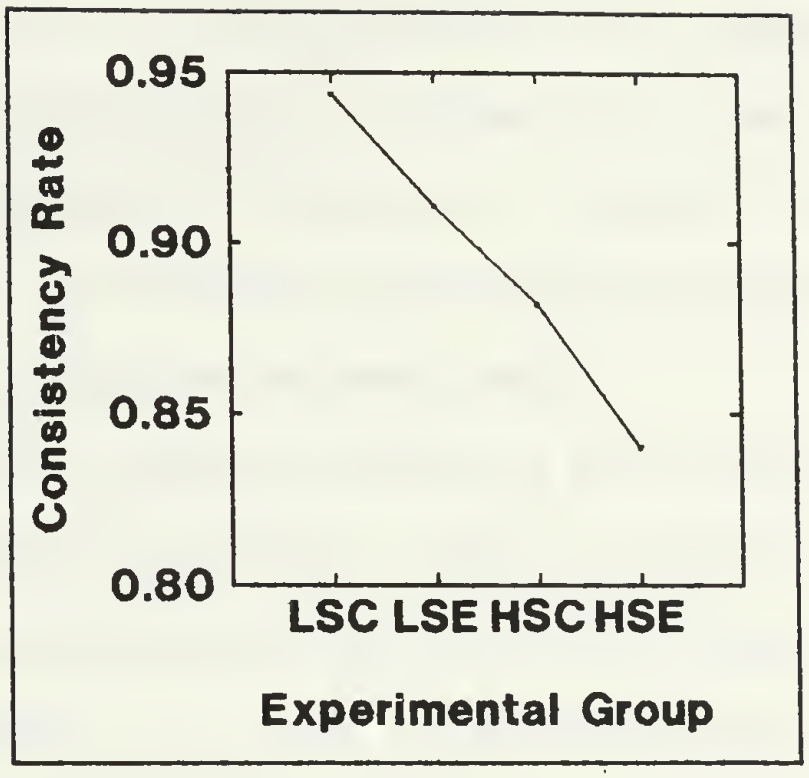

Figure 3. Mean me/not me consistency rate by subgroups: low-splitting control (LSD), low-splitting experimental (LSE), high-splitting control (HSC), and high-splitting experimental (HSE).

I expected consistency rates to vary according to the degree to which a trait adjective was rated as like or unlike the self. As Figure 4, Figure 5, and Figure 6 illustrate, observed consistency was lower for all subgroups in the case of traits rated moderately, rather than extremely. This is what prior schema research would predict; the more schematic or antischematic a dimension, the more consistent one's judgments on that dimension should be.

Table 3 presents the mean consistency rates of me/not me judgments of trait adjectives by the four experimental groups for adjectives originally rated antischematically unlike the self (ratings of 1-3), schematically like the

self (ratings of 7-9), or aschematically (ratings of 4-6). In addition to reflecting a tendency for lowest consistency to occur in judgments of aschematic traits, the table shows that the range of the means differed most for adjectives rated aschematically or schematically.

However *t*-tests failed to find any of the group means significantly different from any other. In addition to the problem of relatively small *N*, noted above, the power of the *t*-test suffers in a case like this by the loss of variance information which occurs when a summary dependent variable is constructed from other variables in the raw data, especially when the latter represent repeated measures. In this study subjects made two me/not me judgments about each trait adjective and the consistency rate refers only to the average congruence of judgments involving all 36 trait adjective stimuli.

Some of the stimuli are of more theoretical interest than others, given the earlier discussions of borderline splitting and the tendency for borderlines to form good and bad self-schemata. Wondering if important intergroup differences in the consistency of responses to the terms *good* and *bad* were being swamped by variance in the overall consistency rate attributable to other terms, I examined the good and bad responses in isolation. It turns out that subjects with low Gerson scale scores were perfectly consistent in their responses to the terms *good* and *bad*.

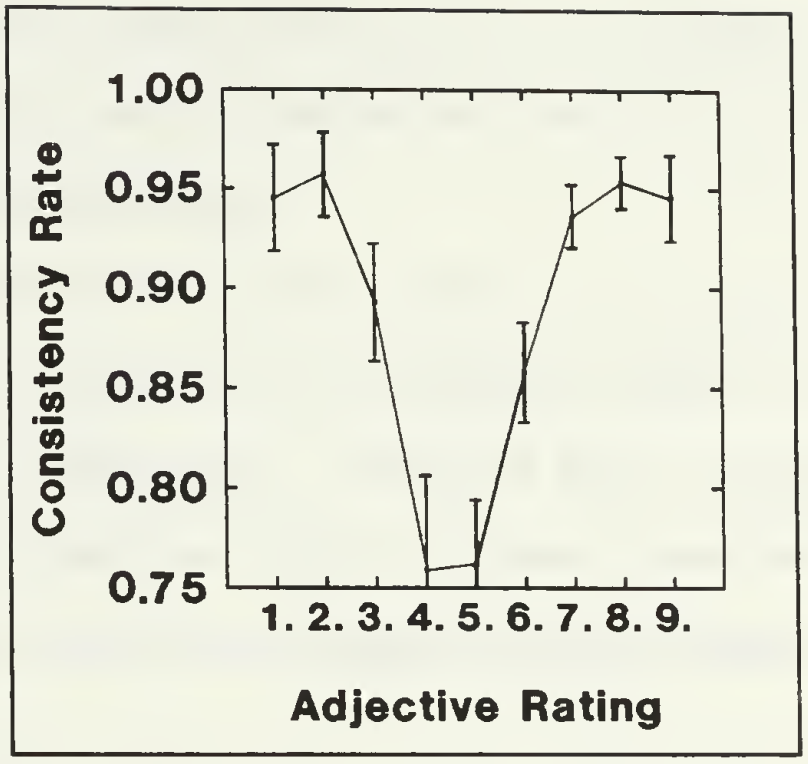

Figure 4. Mean consistency rate for all subjects by rating of adjective's self-descriptiveness.

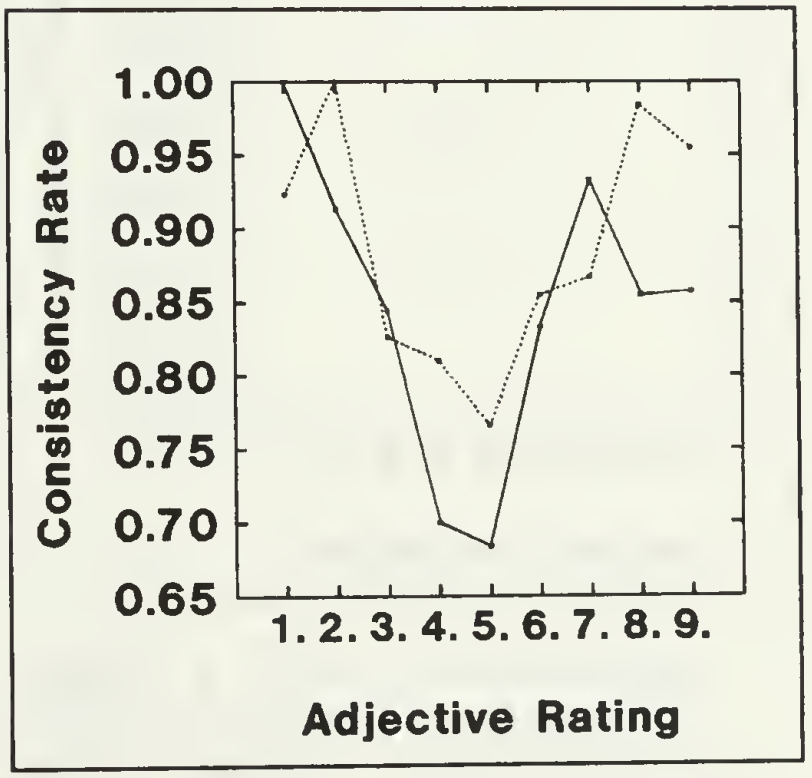

Figure 5. Mean consistency rate of me/not me judgments for frequent splitters by trait ratings. (Solid line is for experimental condition, dashed line for control condition.)

However, the consistency rates for the frequent splitting subjects, in the control and experimental conditions, were .84 and .83, respectively.

It was obvious from mere observation of the group means that the experimental condition could not have had much impact on consistency, because both the infrequent splitting subgroups had identical mean consistency rates and the two frequent splitting subgroups had means that were very close.

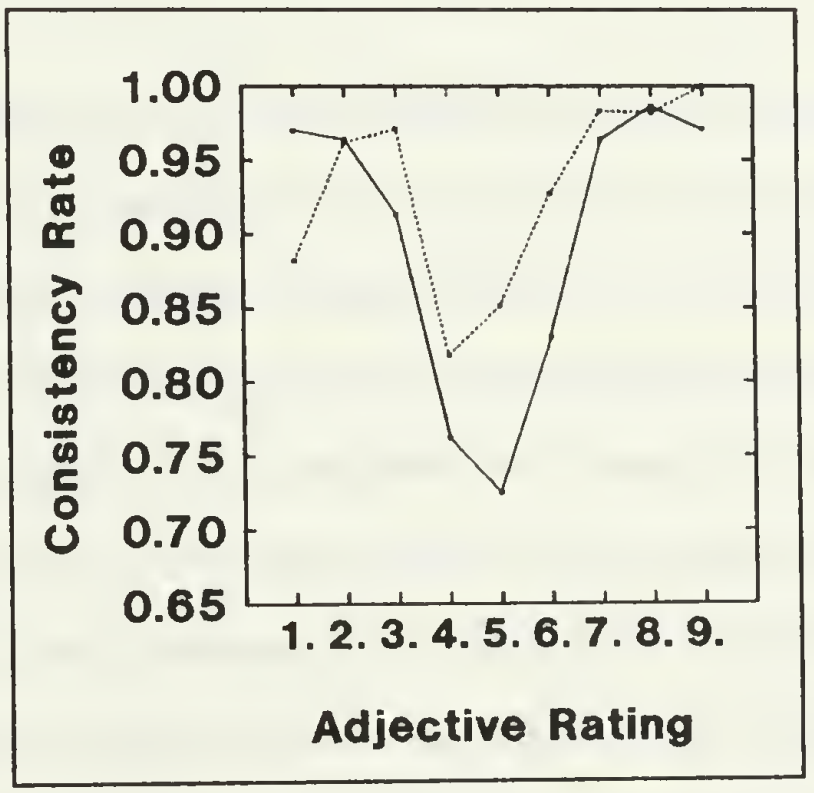

Figure 6. Mean consistency rate of me/not me judgments for infrequent splitters by trait ratings. (Solid line is for experimental condition, dashed line for control condition.)

Table 3. Mean consistency rate of me/not me judgments for the four experimental groups, by degree to which a trait adjective was originally rated as like the self.

Adjective Rating	Groups LSD	LSE	HSC	HSE	Largest Difference
Schematically Unlike the Self (1-3)	.94	.95	.94	.88	.07
Aschematically (4-6)	.89	.79	.82	.78	.11
Schematically Like the Self	.99	.97	.94	.87	.12

Due to the lack of variance in two of the four subject groups I could not perform an *Analysis of Variance* (ANOVA). Therefore, I used linear regression to test the effect of propensity for splitting on the consistency rate for the terms *good* and *bad*. In a regression equation (as opposed to an ANOVA equation) it made sense to use the actual Gerson scale score as the independent variable, rather than the less informative dichotomous variable for designating the Gerson score as high or low. The results of the regression are presented in Table 4.[4]

These results provide some support for a restricted version of hypothesis 1, suggesting that high Gerson scorers

[4] A *t*-test was conducted to test the hypothesis that μ is different for low Gerson scorers and high Gerson scorers, producing similar results: $t(18) = -2.0497$, $p \approx .025$.

Table 4. Results of regression of consistency of me/not me judgments for the words *good* and *bad* on Gerson splitting scale score.

Variable	B	SE	Beta	T	One Tail p
Constant	.97	.043	.000	22.804	.000
Gerson Score	-.003	.001	-.326	-2.101	.021

N=39 R=.39 R^2=.107

tend to be somewhat more likely than low scorers to respond inconsistently to the global traits *good* and *bad*.

In view of the fact that the R^2 of .107 is not especially large, two observations are important. First, the sample for this study represents a nonclinical population. The prediction of greater inconsistency of me/not me judgments involving the *good* and *bad* terms was derived, essentially, from the borderline literature. Thus, there is reason to expect one would achieve a higher R^2 and lower p if the regression was performed using a clinical subsample.

Second, high and low Gerson scorers differ significantly in consistency rates of me/not me judgments of the terms *good* and *bad* but the difference in consistency rate for all trait terms is not significant. This provides additional evidence that although the inconsistencies in *good* and *bad* responses were few in number, they were meaningful, and not simply due to random error.

Hypotheses 3a and 3b

Hypothesis 3a states that subjects with high Gerson scale scores will deviate further than subjects with low Gerson scores from the midpoint value on the 9-point scale on which they are asked to rate the self-descriptiveness of the trait term *good*.

Hypothesis 3b states that subjects with high Gerson scale scores will deviate further than subjects with low Gerson scores from the midpoint value on the 9-point scale on which they are asked to rate the self-descriptiveness of the trait term *bad*.

To test these hypotheses I measured subjects' deviation from the midpoint of the 9-point rating scale on which they indicated the degree to which they saw themselves as like or unlike a trait. An overall deviation score (D) was calculated simply as: $D = \Sigma\ |x-5|$, where x equals the rating a subject gave a particular trait adjective.

Because D measures deviation from the midpoint in a linear manner, three instances in which a subject deviates by one from the midpoint contribute as much to D as a single response deviating by three points. Thus, D does not necessarily capture occasional deviation of large magnitude. To accentuate the most schematic responses, I constructed another measure: $D_s = \Sigma\ (x-5)^2$, with the subscripted s in D_s referring to the squaring of differences.

The direction of results, as reported in Table 5, opposes the predictions of Hypothesis 3a and Hypothesis 3b. Subjects with low Gerson scores had higher *D* and D_s scores than those with high Gerson scores (values for *t* and two tailed *p* are included for reference only).

Table 5 has several interesting features. First, note that the magnitude of mean difference for D_s, under all trait adjectives, is sufficiently large that if I had correctly predicted the direction of the difference it would be statistically significant under the single tailed criterion (*p* would then equal .041 instead of .082). The same is almost true for *D*, where *p* would equal .055 under single tailed criterion, assuming the differences were in the predicted direction. Thus, the results suggest that low Gerson scorers actually rate themselves as more like or more unlike many diverse trait adjectives than high Gerson scorers.

Hypotheses 4a and 4b

Hypothesis 4a states that subjects with high Gerson scale scores will respond more rapidly in making judgments about the trait adjective *good* than subjects with low Gerson scores.

Table 5. Mean values of D and D_a for the low and high Gerson scale score groups.

	Good and *Bad*		All Trait Adjectives	
Group	D	D_a	D	D_a
Low Gerson Scores	2.605	8.026	2.151	6.239
High Gerson Scores	2.263	6.263	1.943	5.239
	t_{36}=1.146 p=.259	t_{36}=1.256 p=.217	t_{37}=1.636 p=.110	t_{37}=1.788 p=.082

Hypothesis 4b states that subjects with high Gerson scale scores will respond more rapidly in making judgments about the trait adjective *bad* than subjects with low Gerson scores.

The test of hypotheses 4a and 4b involved comparison of three measures: 1) the response latency for rating the trait terms on a nine point scale, 2) the response latency for trait terms during the first set of me/not me judgments, and 3) response latencies for the third set of me/not me judgments. Results are presented in Table 6.

Contrary to hypotheses 4a and 4b, the mean reaction time of high Gerson scorers exceeded that of low Gerson scorers in every instance, although never to a statistically significant degree.

Note that both high and low Gerson scorers tended to respond more rapidly to the words *good* and *bad* than other trait adjectives, and that high Gerson scorers tended to respond more slowly in all tests than low Gerson scorers. Given these facts, it made sense to test the possibility that high Gerson scorers responded more rapidly to the words *good* and *bad*, in comparison to the baseline provided by all

Table 6. Mean reaction times (ms) to the words *good* and *bad*, and all trait adjectives, on various rating and judgment tasks.

Variable	Low Gerson Scores	High Gerson Scores
Nine Point Rating of all Trait Adjectives	2590	3186
Nine Point Rating of *good* and *bad*	2491	3052
First Me/Not Me Judgment of all Trait Adjectives	1022	1086
First Me/Not Me Judgment of *good* and *bad*	850	988
Second Me/Not Me Judgment of all Trait Adjectives	938	1042
Second Me/Not Me Judgment of *good* and *bad*	843	972

Table 7. Ratios formed by dividing the mean reaction times for the terms *good* and *bad* by the mean reaction times for all other trait adjectives.

Task	Low Gerson Scorers	High Gerson Scorers
Nine Point Rating	.961	.958
First Me/Not Me Judgment	.832	.910
Second Me/Not Me Judgment	.899	.933

other trait adjectives, than low Gerson scorers. Table 7 provides the observed ratios of average reaction time for the words *good* and *bad* in comparison to reaction times for other adjectives.

None of the differences in reaction ratios are statistically significant, or approach significance, and in two of the three cases high Gerson scorers in this study had a higher observed ratio than low Gerson scorers. This unexpected result could be due to the previously noted fact that high Gerson scorers in my sample rated the words *good* and *bad* less extremely than low Gerson scorers. As shown in Figure 7 and Figure 8, reaction times for all adjectives on the me/not me judgments tended to be highest when terms were rated near the middle of the nine point scale, and lowest when the terms were rated extremely, replicating Kuiper's (1981) "inverted-U" finding.

The question this raised was whether high Gerson scorers might have actually reacted more quickly to the terms *good* and *bad* than low Gerson scorers, controlling for extremity of initial adjective rating. Table 8 sheds light on this question. It compares frequent and infrequent splitters in terms of the ratio formed by dividing the reaction time for the first me/not me judgment of the words *good* and *bad* by the average reaction time for all adjectives on the same task, with results organized according to whether subjects initially rated the words *good* and *bad* schematically or antischematically (deviating by 3 or 4 points from the midpoint on the nine point scale) or aschematically.

As Table 8 indicates, the infrequent splitters had lower observed ratios in all categories, except for the word *good* when it was schematically rated. None of the differences of means between frequent and infrequent splitters, within categories of schematicism and word, were statistically significant.

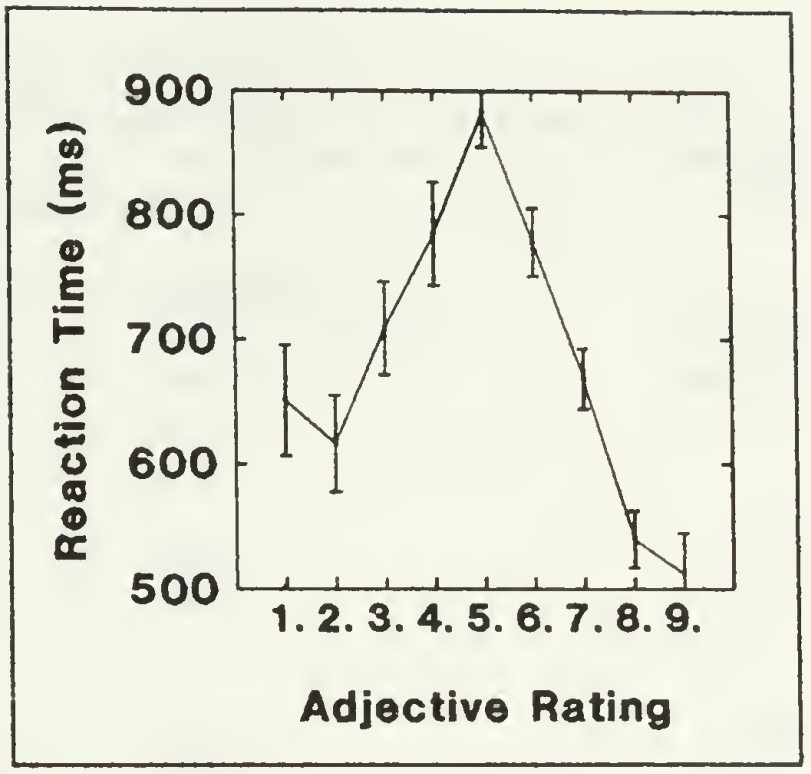

Figure 7. Mean reaction time on the first me/not me judgment task, for all subjects, by rating of adjective's descriptiveness of the self.

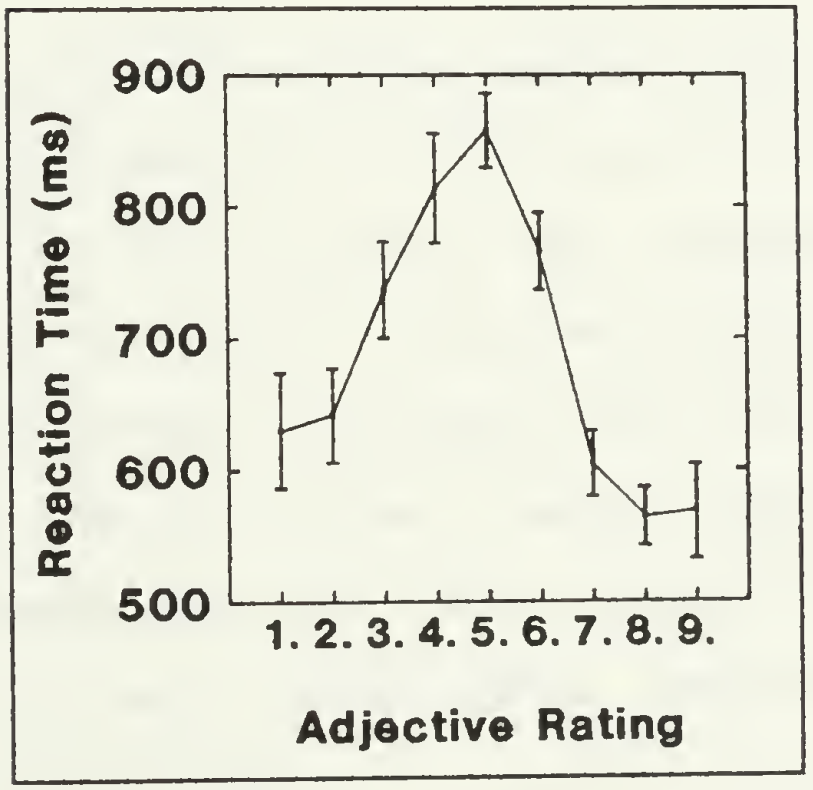

Figure 8. Mean reaction time on the second me/not me judgment task, for all subjects, by rating of adjective's descriptiveness of the self.

Table 8. Ratios formed by dividing the mean reaction times for the words *good* and *bad* by the mean reaction time for all trait adjectives on the first me/not me judgment task, by degree of schematicism on the good and bad dimensions.

Word	Gerson Score	Schematic or AntiSchematic Rating	*n*	Aschematic Rating	*n*
good					
	Low	.688	12	.601	7
	High	.652	7	.815	11
bad					
	Low	.641	10	.735	9
	High	.834	9	1.026	10

Discussion

Hypothesis 1 predicted that subjects with high Gerson scale scores would show more inconsistency in me/not me judgments than low Gerson scorers. This received limited support from the fact that the observed differences in mean consistency rate closely approached statistical significance when responses under 100 ms were removed from analysis. Also, the regression of consistency rate for the words *good* and *bad* on actual Gerson scale scores, indicated that consistency of self-judgment on the good and bad dimensions declines as tendency to split increases.

The results did not support Hypothesis 2, which predicted that subjects in the experimental condition would have lower consistency rates than subjects in the control

condition. This may be partly due to the fact that the dichotomous nature of the control variable could not be circumvented, in the same way that the dichotomous assignment of subjects to low or high Gerson scale score status was overcome in the regression equation when actual scale scores were used as the independent variable.

Inferences to a larger population aside for a moment, the mean consistency rates observed in the four subsamples (Figure 3) did follow the pattern predicted by Hypothesis 1 and Hypothesis 2. It appears that the large variance in the consistency variable reduced statistical power to the point where even big observed differences of mean provided little basis for inference. Thus, there is some reason to remain optimistic that the null hypotheses associated with Hypothesis 1 and Hypothesis 2 might be rejected in a more powerful experiment.

On the other hand, observed differences of means failed to offer any support at all for Hypothesis 3a or Hypothesis 3b. Thus, there is no evidence that high Gerson scores in the population is associated with a greater likelihood of rating the adjectives *good* and *bad* self-schematically or antischematically. However, because there are other features of schematicism than those measured by simple self-rating on a dimension (Fiske and Taylor, 1991), the test of Hypothesis 3 does not constitute a complete test of the general hypothesis that frequent splitters are more schematic or

antischematic on the *good* and *bad* dimensions. This deficiency of Study 1 is directly addressed in Study 2, where all three of the Fiske and Taylor (1991) criteria are used in assessing schematicism.

Hypothesis 4a and Hypothesis 4b received no support from the present results. There is no evidence that high Gerson scores in the population are associated with lower reaction latencies in making conscious me/not me judgments about the terms *good* or *bad*.

The most elegant interpretation of these results is that the high Gerson scorers were no more schematic or antischematic on the good/bad dimension than low Gerson scorers. Thus, they did not rate the words *good* or *bad* more extremely self-descriptive and did not respond to them more rapidly. The lower consistency rate in me/not me judgments, in this view might be due to something other than splitting, such as a more muddled sense of self, or a greater tendency to respond to subtle demand features of the experiment.

Given the weak power of the present experiment, however, and the incomplete measurement of schematicism, it seems premature to draw any firm conclusions. I critiqued Study 1 with the goal of improving power in a follow-up experiment to retest the original hypotheses and a few additional ones.

One problem, the large variance between the individual consistency rates of subjects, can only be remedied to the

extent that this variance includes error. The part of the sample variance which reflects real variance in the population is inevitable in any sample, but sample variance due to measurement errors could potentially be reduced.

In Study 1 a subject responded to each adjective once in each me/not me judgment trial. Thus, a single response error could make what would have been a consistent response inconsistent, or vice versa. If, in contrast, a subject responded several times to each adjective, simple response errors would contribute less to variance in overall consistency rate, or consistency rate for a given adjective.

Another way to tame the large variances in consistency rates would be to sample more subjects, taking advantage of the fact that *N* and *n* serve in the denominator of the formula for standard error. Unfortunately, practical considerations made 40 subjects the maximum number that could be included in an experiment like this one.

Study 1 did not include a manipulation check other than some discussion during debriefing about the impact of the music upon the subjects' mood. Some subjects reported having entered the mood of the music and described their feelings about the music in some detail, but many subjects reported that they had experienced little mood change. Unfortunately, this information was not systematically gathered.

Interestingly, a number of subjects reported being sufficiently alarmed by the language of the informed consent

form that they later believed they had resisted mood induction without really intending to do so.

Thus, Study 1 is open to the criticism that it failed in some cases to induce the intended moods and that it is now impossible to sort out where mood induction worked and where it did not. This criticism is directly relevant to interpretations of data relating to Hypothesis 1 and Hypothesis 2 because the dependent variable for the test of these hypotheses (consistency rate) involves a measure taken after the second variable mood induction. The tests of Hypothesis 3a and Hypothesis 3b, in contrast, involved data taken prior to any mood induction, and two of the three measures used to test Hypothesis 4a and Hypothesis 4b were taken before the variable mood induction.

Another problem with this experiment, relevant only to the test of Hypothesis 4a and Hypothesis 4b, involves the reliability of time measurements using the internal timer of the PC. The technical issues are detailed in Appendix A. Here it should be noted that although timing error was a major source of variance in all the timing data it could at least be treated as random error. Thus, even if timing error lowered the power of statistical tests it did not bias them.

Another criticism of Study 1 deals with the fact that it only employs transparent measures of self-judgment. This raises important issues. First, it increases the likelihood that subjects understand, on some level, what is expected of

them in the study. Some subjects may tend to comply with such an implicit demand, even as others may tend to go against the grain of expectation.

Because the judgment tasks are all aimed at subjects' conscious sense of self, there is also a likelihood that simple desire to maintain a consistent self-image entered the equation for some subjects. Consistency of self-image is apparently important for most people, in life as in experimental situations (Fiske & Taylor, 1991, pp. 218-223). There is no *a priori* reason to think that the need for consistent self-image is any less significant to borderlines or other frequent splitting persons, except that with frequent splitters inconsistency is more likely to be dealt with via splitting instead of the less dramatic methods employed by people who do not often split. In any case, it would be nice to have measures of current sense of self which are not so vulnerable to either demand effects or a desire for self-consistency.

The fact that only conscious measures of sense of self were employed also raises an interesting theoretical point. Although a sense of self is undoubtedly a crucial component of many day to day decisions and judgments, it is often located in the cognitive background, not at front center stage as the tasks of study 1 demand.

When one has developed a fairly inclusive, coherent, singular sense of self, calling it directly into

consciousness may make little difference to its information processing effects. But what if one harbors two split self-images? In this case is it not possible that by directly bringing self-concept to the forefront of consciousness one also increases the chance that inconsistencies between current feedback about the self and the activated side of self-concept will arouse partial awareness of the opposite sense of self?

From a psychoanalytic perspective, the bringing to consciousness of the contradiction between the activated and deactivated self-representations represents a failure of splitting, necessitating further defense such as denial, isolation of affect, etc. One presumes that any such defensive operations use up limited cognitive capacity to a certain degree. This would impair any other mental processing occurring at the same time (such as those demanded by the judgment tasks of Study 1) with slowing of reaction a natural secondary effect.

Thus, an additional reason for desiring measures of an activated sense of self which does not directly call the qualities of that sense of self into consciousness, is that unconscious measures of self-concept are less likely to call into play a series of other defenses which may complicate the results.

A final criticism of Study 1 deals with the Gerson scale, which has been little studied since its introduction

in 1984. Although Glassman (1986) examined its formal psychometric qualities, shortening it by half as a result of his findings, the scale has received little testing to establish its validity. Before Gerson's scale is used again it is desirable to see if it is associated in ways we would expect of a splitting measure with other variables contained in the prescreening data base.

Summary

In this chapter I described the procedures and hypotheses associated with Study 1, presented results from statistical tests of the hypotheses, and drew conclusions regarding the validity of the hypotheses.

CHAPTER 3

STUDY TWO

Study 2 was organized with three goals in mind: 1) to test the validity of the Gerson scale, 2) to retest several of the hypotheses from Study 1, or variations of them, with improved statistical power, and 3) to test additional hypotheses involving unconscious aspects of self-concept in frequent and infrequent splitters.

Methods

This section describes the procedures of Study 2.

Sampling

Study 2 used the same prescreening and sampling procedures used previously in Study 1. The final sample for Study 2 included 21 subjects with low Gerson scale scores (3 males, 18 females) and 20 subjects with high Gerson scale scores (3 males, 17 females). Low Gerson scores ranged from 1 to 4, and high scores ranged from 20 to 25.

Study Procedures

A detailed description of procedures used in Study 2 is available in Appendix B. Only a general description of procedures is presented here.

My assistants and I conducted Study 2 in a small experimental room at the University of Massachusetts at Amherst. The room contained a desk, two chairs, and a single IBM-compatible PC .

An assistant or I met with each subject for the hour the study procedures required. We welcomed students to the study and read the informed consent form to them. Once the form was signed, the study began.

We led subjects to falsely believe that procedures they would undergo were intended to aid the development of a new type of cognitive test. This deception was used to explain the presence of speed and accuracy tests for which subjects would receive mood-inducing performance feedback, as outlined below and described in detail in Appendix B.

After providing some demographic data in response to prompts on the computer monitor, subjects were prompted by computer to describe their current mood using a maximum of three adjectives. Then the computer asked subjects to provide two adjectives which they considered extremely self-descriptive and important. Subjects were also asked to provide antonyms for these two words, yielding a total of

two words labelling idiographic schematic traits and two words labelling idiographic antischematic traits.

The computer then presented an increasingly difficult reverse digit recall test and gave all subjects rigged feedback indicating very poor performance in relation to peers. This constituted the dysthymic mood induction. The computer probed for mood after the induction was completed.

Next the computer directed subjects to watch a bracketed area of the monitor and to observe masking symbols that were flashed there. Subjects were then asked to judge whether a succession of trait terms subliminally presented in the masked location were better defined as *me* or *not me*. Registering this judgment required subjects to press an appropriate keyboard key as rapidly as possible after they formed an "intuition" about which response (*me* or *not me*) was best. Subjects were tested 60 times, with each of 10 adjectives being presented six times. The 10 adjectives included the terms *good*, *bad*, *loving*, *hateful*, *passive*, and *assertive*, as well as the four idiographic terms provided by the particular subject.

Next the computer randomly presented the adjectives, in a fully perceptible way, six times each. Subjects were again asked to make me/not me judgments as rapidly as possible.

We then showed the subjects card 9GF of the Thematic Apperception Test (TAT) and asked them to rate the two characters on the card on the 10 trait dimensions.

Next we attempted a happy mood induction by having subjects respond via computer to a sham test of their ability to distinguish a particular symbol, in a fast and accurate manner, from other symbols presented rapidly on the monitor. Subjects then received rigged feedback to the effect that they had performed very well on this test. The tasks involving subliminal and perceptible presentation of trait adjectives were repeated and mood was queried. Subjects were then debriefed and dismissed.

On the Lack of a Control Group

Study 2 contained no control group. All subjects received the same experimental manipulations (the performance feedback) in the same order. This creates problems in interpreting results, of course, and the decision not to use a control group was not made lightly.

The labor intensive experimental procedure, which required the presence of the experimenter for one hour with each subject, made it impossible to run more than 40 subjects in the time that was available. Because half had to be frequent splitters, and half infrequent splitters, the projected *n* was 20 for each subgroup. Any further division of the subgroups, for example, into control and experimental groups, would have reduced the n of subgroups to 10, as in Study 1. In Study 1 statistical power was poor, in major

part because of the low numbers of subjects in identifiable subgroups. Power simply had to be improved in Study 2, and because most of the major hypotheses dealt with differences between frequent and infrequent splitters, and not the effect of manipulations, per se, eliminating the control group seemed reasonable.

The reader should keep in mind, however, that because of the lack of a control group in Study 2, the idea that manipulations caused any given effects must always be taken with a grain of salt. Inferences about differences between frequent and infrequent splitters, under the sort of conditions subjects experienced in this study, stand on more solid ground.

Scoring the Mood Data

An independent evaluator, unaware of the purpose of the mood data, compared the blocks of mood responses from immediately before and after the mood inductions, and judged whether mood was more or less euthymic in the second block. The evaluator examined responses of the 34 subjects who showed some response variation across the four mood enquiries. Table 9 summarizes the results.

It is especially noteworthy that although 23 of the 34 subjects reported less euthymic mood after receiving negative performance feedback, only 8 reported a more

euthymic mood after the positive feedback.[5] This suggests that although the first mood induction was generally effective the second one may not have been.

Close analysis of subjects' later mood responses, and information from the debriefings with subjects after the experiment, suggest that by the time of the second mood induction most subjects were tired (and often irritable) as a result of paying close attention to repetitious procedures. Any positive effect of favorable performance feedback seems to have been diluted, and even negated, by the fatigue which had by that point set in. This conclusion is supported by the fact that results from the second half of the experiment are negligible, reflecting what appears to be fairly random responding to the major experimental tasks. Accordingly, I decided not to report results from the second half of Study 2 in the present report.

Instead of relying on an omnibus measure of mood change (euthymia/dysthymia) I grouped mood responses under four categories: anger, boredom, anxiety, and sadness. For example, responses such as "annoyed", "mad", and "angry" were considered indicative of anger. Appendix C lists other terms categorized under each heading.

---

[5] Only mood change, and not absolute euthymia is at issue here. A subject could be judged as less euthymic from Time 1 to Time 2 and yet have responded in both instances in a way which suggested general euthymia.

Table 9. Number of subjects judged more or less euthymic after the mood inductions, by overall response pattern.

	More or Less Euthymic ?	
n	After Negative Feedback	After Positive Feedback
1	+	+
3	+	-
6	-	+
10	-	-
1	no change	+
1	no change	-
1	+	no change
7	-	no change
4	no change	no change
7	(all mood responses identical)[a]	
Total -	23	14
Total +	5	8
Total no change	6	12

[a]These subjects responded identically to the four mood queries. In contrast, the 4 subjects categorized as "no change" under both mood inductions respond differently, but these differences were not judged more or less euthymic.

Four dichotomous variables were established to indicate whether or not a subject expressed anger, sadness, anxiety, or boredom in the mood inquiries temporally associated with the first block of experimental tasks. Each variable was coded zero if a subject gave no response indicative of the respective affect, or a response indicative of its opposite. Thus, a response of "happy" resulted in a zero code under the variable for sadness. A value of one, on the other hand, was coded if a subject indicated "sadness" in some way.

Merging Prescreening and Study Data

The psychology department provided access to the SPSS system file containing extensive data on each of the hundreds of undergraduates prescreened at the beginning of the semester, including the 41 participants in the present study. I generated a new SPSS file containing only the data for these 41 students, sorted it by social security number, and imported into it the data from Study 2.

This gave me access to more than 1,000 variables generated either by the present research or by other researchers during the prescreening administration. This was fortuitous in that it made it possible to compare Gerson scale scores with a borderline personality disorder scale, and prescreening scores on the Beck Depression Inventory, the Spielberg Anxiety Scale, and other interesting psychological measures.

Validating the Gerson Splitting Scale

The present study's reliance on Gerson's splitting scale cannot be overemphasized. The scale was integral to the disproportional stratified method of sampling, and its validity is thus of prime importance.

I have already discussed Glassman's (1986) psychometric validation of the Gerson scale. In Study 2 I tested its

construct validity by seeing if it correlated with variables one would expect to be associated with splitting.

Because people with borderline personality disorder comprise one of the largest groups engaging in frequent splitting, any scale which purports to assess predisposition to splitting should correlate very highly with a valid borderline scale. Thus, I predicted that the Gerson scale would correlate positively with ten items most indicative of borderline personality in the Millon Clinical Multiaxial Inventory (Millon, 1987), also administered during prescreening. The observed correlation among all prescreened subjects of .5064 ($p<.01$) supports the prediction.

I also predicted that Gerson scale score would correlate positively with Beck Depression Inventory (BDI) score and Spielberger Trait Anxiety Inventory (STAI) score (Spielberger, Gorsuch, & Lushene, 1970). These predictions were based on the psychoanalytic view of splitting as an indicator of underdeveloped defensive ability, overreliance on "primitive" splitting-based defenses, and underuse of more efficient defenses based on repression (Blanck & Blanck, 1974). Less efficient defenses should correlate with greater psychological distress, including depression and anxiety (Kernberg, 1975).

The observed correlations between Gerson scale score and BDI (.5189, $p<.01$) and STAI (.6458, $p<.01$) for all prescreened subjects support these predictions.

As Lerner and Lerner (1982) and Lerner (1991) noted, intrapsychic splitting can manifest as differential projections onto two figural percepts. Thus, I expected the frequent splitting experimental group to differentiate the two TAT characters on the 6 standard and 4 idiographic trait terms. The results presented in Table 10 support these predictions. Differentiation of TAT characters (x) was computed quite simply: $x=\Sigma\ |(TAT_1 - TAT_2)|$, where TAT_1 refers to the numerical score a subject's first named character received on a given trait dimension, and TAT_2 refers to the score the second named character received on the same trait dimension.

Table 10. Results of *t*-tests comparing mean differentiation of TAT characters by frequent and infrequent splitters.

Variables	mean	n	sd (se)	t (df)	p
Gerson Group					
Standard TAT					
Infrequent Splitters	6.095	21	4.76 (1.04)	-1.95 (38.23)	.029
Frequent Splitters	9.150	20	5.22 (1.17)		
Idiographic TAT					
Infrequent Splitters	4.000	21	1.87 (0.41)	-1.76 (26.42)	.045
Frequent Splitters	5.750	20	4.06 (0.91)		

The predicted covariance of Gerson Scale scores with other indicators of splitting was fully observed, providing construct validation for the Gerson Scale.

Retesting Hypotheses from Study One

I was able to retest several of the hypotheses of Study 1, although due to differences in experimental setup these retests fall short of true replications.

As in Study 1, I expected frequent splitters to show greater inconsistency than infrequent splitters in their me/not me judgments of fully perceptible trait adjectives following negative performance feedback versus positive performance feedback.

In Study 1 subjects responded once following a happy mood induction and once following either a repeat of the happy mood induction or a sad mood induction. Consistency of response was simple to measure; either the two responses matched or they didn't.

In Study 2, however, subjects responded six times to each perceptibly presented trait term after the negative feedback and six more times after the positive feedback. Allowing for some response error, I took the majority response to a trait term, within a given block of trials, to represent "true" response. Consistency could then be measured as it was in Study 1.

The mean consistency rate for all trait terms was 1.0 for infrequent splitters but only .96 for frequent splitters. This difference of means was quite close to statistical significance ($t(19)=1.71$, $p<.052$).

This result is consistent with those for Hypothesis 1 in Study 1. The raw data from the two studies cannot be combined directly because the data were derived using different procedures. Nevertheless, the odds of achieving respective *p* values of .065 and .052, if the general underlying hypothesis is incorrect, must be considerably less than either *p* value because the odds of getting two results is the product of the individual probabilities, in this case .003. Thus, I consider the underlying hypothesis strongly supported by the two studies: frequent splitters are more prone to make inconsistent me/not me judgments under conditions of contrasting mood inductions.

Strong evidence from Study 1 persuaded me I was wrong to expect frequent splitters to rate trait adjectives such as *good*, or *bad*, more extremely like or unlike the self than infrequent splitters. In fact, based on results from Study 1, I suspected the opposite was true. If so, I saw no reason to dismiss the equally counterintuitive prediction that infrequent splitters are more schematic on the dimensions *good* and *bad* than frequent splitters, according to the criteria for schematicism set out by Fiske and Taylor (1991).

To test this revised prediction I calculated a schematicism score for each trait term according to the formula

$$x = (HM_{ii} - HM_{sm}) (HI_{ii}),$$

where HM_{ii} equals a subject's rating of self-identification with the term in question, HM_{sm} equals a subject's rating of self-identification with the antonym of the primary trait term, and HI equals the subject's estimation of how important the trait in question is in a person.

Because HM_{ii}, HM_{sm}, and HI_{ii} all took values from 1 to 5 the composite measure of schematicism had a possible range of -20 to 20. A score of 20 represents prototypical schematicism involving maximum endorsement of the primary trait term as self-descriptive, maximum disavowal of the self-descriptiveness of the antonym, and maximum rating of the importance of the trait. A score of -20 reflects prototypical antischematicism, and a score of zero, achievable only when the primary trait term and its antonym are rated equally self-applicable, indicates aschematicism.

As shown in Table II subjects usually rated idiographic trait terms more schematically than standard trait terms--hardly a surprising finding. More interestingly, it also indicates that *loving* was rated much more schematically than *good*, in general, and that the terms *assertive* and *passive*

were both rated fairly aschematically. Antonyms for the idiographic schematic terms, and for the standard terms *loving*, *good*, and *assertive* are almost without exception ranked in inverse order, with negative schematicism scores.

Table 11. Trait terms ranked in descending order by mean schematicism for all subjects.

Term	Mean Schematicism Score
First idiographic schematic term	14.22
Second idiographic schematic term	12.61
loving	12.27
good	7.80
Assertive	2.68
Passive	-2.93
bad	-7.54
Antonym for 2nd idiographic term	-11.49
hateful	-12.07
Antonym for 1st idiographic term	-13.76

The predominance of idiographic terms at the extremes and the nearly perfect symmetry of the scores suggest that schematicism was measured here in a valid manner. Additional support for the validity of this measure of schematicism comes from nonlinear regression of mean reaction time for the me/not me judgment of perceptibly and subliminally presented trait terms on mean schematicism scores for the various terms. The best-fitting regression curve resembled Kuiper's (1981) classic "inverted-U", as reflected in Figure 9 and Figure 10.

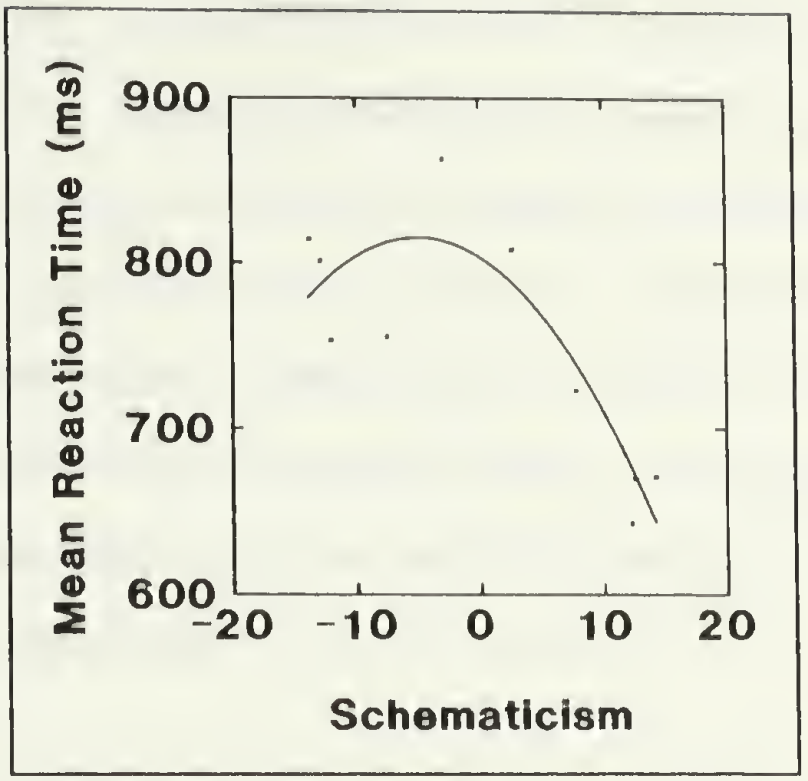

Figure 9 Mean reaction time in milliseconds (ms), for the 10 stimulus adjectives, presented perceptibly for the first time, by mean schematicism score.

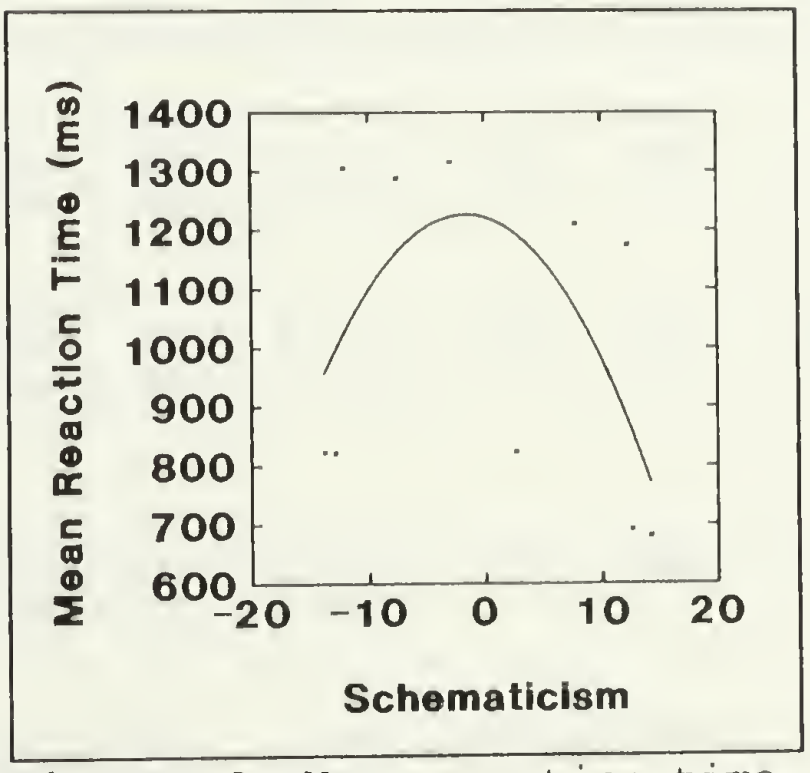

Figure 10. Mean reaction time in milliseconds (ms), for the 10 stimulus adjectives, presented subliminally for the first time, by mean schematicism score.

As predicted, and as reflected in Table 12, infrequent splitters had more extreme observed schematicism scores (i.e., further from zero) than frequent splitters for the words *good*, *bad*, *loving*, and *hateful*. (The difference of means is statistically significant for *good*, *loving*, and *hateful*, but is not quite significant in the case of *bad*.) Differences of means were not significant for idiographic terms or the words *assertive* and *passive*.

Table 12. Results of *t*-tests comparing the mean schematicism scores for frequent and infrequent splitters, for the words *good*, *bad*, *loving*, and *hateful*.

Trait Term / Splitting Group	mean	n	sd	t	df	Single Tail p
good						
Infrequent	9.429	21	4.52	1.81	32	.04
Frequent	6.100	20	1.55			
bad						
Infrequent	-8.952	21	4.96	-1.51	34	.07
Frequent	-6.050	20	7.11			
loving						
Infrequent	15.333	21	4.75	3.71	36	.000
Frequent	9.050	20	6.00			
hateful						
Infrequent	-15.333	21	4.74	-4.02	37	.000
Frequent	-8.650	20	5.82			

Note. Separate estimates of variance used in computing t. Degrees of freedom value rounded to nearest integer.

The last aspect of Study 1 which I retested in Study 2 involved reaction time. Results of Study 1 did not confirm the original prediction that frequent splitters would enjoy a speed advantage in making me/not me judgments when the words *good* or *bad* were presented perceptibly. Despite the fact that the data of Study 2 allowed somewhat more sophisticated testing, results still failed to confirm the original prediction.

Once again the lack of reaction time results may be due to the sort of timing "noise" discussed in Appendix A. This said, it is still worth noting that although frequent splitters rated themselves significantly less schematically on the terms *good*, *loving*, and *hateful*, (contrary to original expectations), they did not also suffer the speed disadvantage one would expect to flow from decreased schematicism. This suggests the interesting possibility that schematicism, as measured by Fiske and Taylor (1991), does not covary with less conscious measures of schematicism in the same way it does for infrequent splitters.

New Hypotheses in Study Two

The data base from Study 2 is richer than that from Study 1 thanks to the addition of tasks involving subliminal presentation of trait terms, the TAT indicator of splitting, etc. This permitted me to test several new hypotheses

dealing with response differences between frequent and infrequent splitters when stimuli are presented subliminally.

Hypothesis 5a, 5b, 5c, and 5d

Even though frequent splitters have not been shown to respond more rapidly to the terms *good*, *bad*, *loving*, or *hateful* when these are perceptibly presented, I predicted greater speed on the subliminal tasks. This prediction was based on speculation that activation of a self-schema would have a more pure effect on processing efficiency when uncomplicated by conscious considerations involved in any self-judgment.

For example, subliminal presentation of stimuli would presumably minimize demand effects, effects based on a desire to appear consistent, and so forth.

Also, it seemed possible that when frequent splitting subjects were asked to make conscious self-judgments on schematic dimensions they became aware not only of the currently activated side of the split but also the muted, but not entirely forgotten deactivated side of the split. If such a process complicated the self-judgment it would also slow reaction time. Subliminal presentation, I speculated, would avoid any such complication. Thus, frequent splitters could conceivably show faster reaction times on the

subliminal tasks, relative to their baseline reaction speed, than infrequent splitters, even if infrequent splitters showed faster times when stimuli were presented perceptibly.

For these reasons, Hypothesis 5a states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *good*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5b states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *bad*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5c states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *loving*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5d states that on the subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *hateful*, relative to their reaction time to the idiographic trait terms.

Hypotheses 6a, 6b, 6c, and 6d

Hypotheses 6a, 6b, 6c, and 6d predict, in general, that frequent splitters will respond with less randomness to the terms *good*, *bad*, *loving*, and *hateful* presented subliminally,

when asked to "guess" whether the consciously unperceived word is *me* or *not me*.

These hypotheses are predicated upon the idea that activation of a self-schema primes a person to respond to stimuli closely associated with the schema. In the case of subliminally presented stimuli, such priming should result in less random judgments of self-descriptiveness. Thus, if a good self-schema has been activated, a subject should be able to respond less randomly when asked to guess if the subliminally presented word *good* is self-descriptive or not.

Nonrandomness would manifest as response directionality--a definite surplus of *me* over not me responses, or vice versa. Using a statistical measure of nonrandom response pattern, a statistic I created and dubbed the *improbability score*, I developed four hypotheses predicting that frequent splitters would respond less randomly to certain stimuli on the subliminal tasks.

Specifically, Hypothesis 6a predicts that frequent splitters will achieve higher improbability scores for the word *good* on the first subliminal task.

Hypothesis 6b predicts that frequent splitters will achieve higher improbability scores for the word *bad* on the first subliminal task.

Hypothesis 6c predicts that frequent splitters will achieve higher improbability scores for the word *loving* on the first subliminal task.

Hypothesis 6d predicts that frequent splitters will achieve higher improbability scores for the word *hateful* on the first subliminal task.

Results

In this section I present results of tests of hypotheses 5a through 6d.

Hypotheses 5a, 5b, 5c, and 5d

Hypothesis 5a states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *good*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5b states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *bad*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5c states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *loving*, relative to their reaction time to the idiographic trait terms.

Hypothesis 5d states that on the first subliminal task the frequent splitters will respond more rapidly than infrequent splitters to the words *hateful*, relative to their reaction time to the idiographic trait terms.

Table 13 lays out the evidence, showing that there is no basis for rejecting the null hypotheses associated with hypotheses 5a or 5b. Contrary to predictions the observed reaction times for the words *good* and *bad*, relative to reaction times for the idiographic terms, were actually slower for frequent than infrequent splitters.

Table 13. Results of *t*-tests comparing ratios formed by dividing mean reaction times for words *good*, *bad*, *loving*, and *hateful* by mean reaction times for idiographic trait terms, by splitting groups, for 1st me/not me judgment of subliminally presented trait adjectives.

					hyp: $\mu_{\text{fs ratio}} < \mu_{\text{is ratio}}$		
Word	Gerson Scale Score	Average RT (ms)	Ratio with Idiographic Traits	*n*	*t*	df	Two Tail *p*
Good	Low	1039	1.528	21	.525	37	.541
	High	1389	1.633	19			
Bad	Low	1155	1.694	21	.656	38	.966
	High	1429	1.702	19			
Loving							
	Low	1099	1.216	21	3.37	23	.002
	High	1272	.908	19			
Hateful							
	Low	1104	1.190	21	.46	38	.325
	High	1547	1.140	19			

Note. Separate estimates of variance used in computing *t*. Degrees of freedom value rounded to nearest integer.

In contrast, as reflected in Table 13, observed reaction times for the words *loving* and *hateful*, relative to

reaction times for the idiographic terms, were faster for frequent than infrequent splitters. However, only in the case of the word *loving* was this difference of speed statistically significant. Thus, the null hypothesis associated with Hypothesis 5c can be rejected, but not the null hypothesis associated with Hypothesis 5d.

Hypotheses 6a, 6b, 6c, and 6d

Hypothesis 6a predicts that frequent splitters will achieve higher improbability scores for the word *good* on the first subliminal task.

Hypothesis 6b predicts that frequent splitters will achieve higher improbability scores for the word *bad* on the first subliminal task.

Hypothesis 6c predicts that frequent splitters will achieve higher improbability scores for the word *good* on the second subliminal task.

Hypothesis 6d predicts that frequent splitters will achieve higher improbability scores for the word *bad* on the second subliminal task.

The construction of the improbability score is complicated, and is fully discussed for interested readers in Appendix D. For present purposes it should suffice to say that this score represents the improbability of achieving the observed pattern of me/not me responses for a given

trait term by pure chance alone, as calculated by the binomial theorem. The higher the improbability score the less likely the observed pattern is due to chance alone.

Table 14 presents the average improbability weights for frequent and infrequent splitters and a *t*-test of hypothesis that $\mu_{fs} \neq \mu_{is}$.

Table 14. Results of *t*-tests comparing mean improbability scores for response patterns to the words *good*, *bad*, *loving*, and *hateful* for frequent and infrequent splitters, for the 1st subliminal task.

Mean Improbability Score		hyp: $\mu_{fs} > \mu_{is}$		
Word	Frequent Splitters	Infrequent Splitters	*t*	df	*p*
Good	14.399	5.726	-1.87	20.98	.032†
Bad	5.106	6.001	.40	32.17	.688‡
Loving	11.675	6.510	-1.20	21	.121†
Hateful	6.055	5.453	-.42	33	.338†

† One-tailed *p*.
‡ Two-tailed *p*.

Only in the case of the word *good* in the first did the mean improbability weights differ sufficiently between frequent and infrequent splitters to reject the null hypothesis that $\mu_{fs} = \mu_{is}$.

A subject's overt schematicism on the trait dimensions is a potentially important mediating variable in determining

nonrandomness of response to any subliminally presented trait term. Having already seen that frequent splitters tended to rate themselves less schematically on the terms *good* and *bad* than infrequent splitters, it seemed likely that the effect of splitting would be enhanced if schematicism was statistically controlled.

Thus, I regressed improbability scores for all terms on schematicism and splitting scale scores. In the case of the words *good* and *loving* the effect of tendency to split was enhanced by controlling schematicism, as reflected in Table 15 and Table 16. This shows that splitting and schematicism were both positively associated with nonrandom responding to the words *good* and *loving* on the first subliminal task.

Table 15. Results of regression of improbability scores for the word *good* in the 1st subliminal task on schematicism on the good dimension and splitting scale score.

Variable	*B*	*SE*	*Beta*	*t*	*p*
$Schematicism_2$	.820	.402	.329	2.038	.049‡
Schematicism	-.000	.039	-.003	-.020	.984‡
Splitting Scale Score	.551	.234	.371	2.359	.012†
Constant	3.246	3.698		.878	.386‡

Total R = .177
Incremental R^2 due to splitting variable, after controlling for schematicism =.124

†One-tailed *p*.
‡ Two-tailed *p*.

Thinking that mood might also mediate the effect of splitting tendency on response to subliminal stimuli, I entered mood variables in the family of regressions of improbability scores for the words *good*, *bad*, *loving*, and *hateful* in *post hoc* analysis. At the same time, I considered the possibility that differentiation of TAT characters, one indicator of splitting, would also mediate the independent effect of tendency to split.

Table 16. Results of regression of improbability scores for the word *loving* in the 1st subliminal task on schematicism on the loving dimension and splitting scale score.

Variable	*B*	*SE*	*Beta*	*t*	*p*
Schematicism,	1.045	.381	.480	2.741	.009‡
Schematicism	.046	.049	.143	.930	.359‡
Splitting Scale Score	.541	.228	.404	2.378	.012†
Constant	.690	3.846		.179	.859‡

Total R = .199
Incremental R^2 due to splitting variable, after controlling for schematicism = .122

† One-tailed *p*.
‡ Two-tailed *p*.

The entire family of regressions included 16 models, because there were four dependent variables (improbability scores for the trait terms *good*, *bad*, *loving*, and *hateful*) and four mood variables (anger, sadness, boredom, and

anxiety). Thus, Bonferroni correction dictated that mood variables achieve a *p* value of less than .003125 (.05/16) to be considered statistically significant. There was only one instance in which this stringent criterion was met: as reflected in Table 17 the independent variable anger was associated with a *p* value of .003 in the regression of improbability score for the word *good*.

Table 17. Results of regression of improbability scores for the word *good* in the 1st subliminal task on schematicism on the good dimension, the mood variable for anger, differentiation of TAT characters, and splitting scale score.

Variable	*B*	*SE*	*Beta*	*t*	*p*
Schematicism¸	.761	.357	.305	2.135	.039‡
Schematicism	-.006	.035	-.023	-.167	.868‡
Anger	24.921	7.775	.439	3.205	.003‡
TAT	-.664	.279	-.341	-2.381	.023‡
Splitting Scale Score	.550	.211	.370	2.608	.006†
Constant	9.873	4.165		2.370	.023‡

Total R =.42
Incremental R^2 due to splitting variable, after controlling for schematicism =.113
Incremental R^2 due to anger variable =.223

† One-tailed *p*.
‡ Two-tailed *p*.

Comparison of Table 15 and Table 17 is interesting because of the dramatic increase in R^2 from .177 to .42, the difference chiefly attributable to the anger variable. Note that the B coefficient associated with anger, like that associated with splitting scale score, is positive, indicating that both tendency to split and anger correlated independently and positively with nonrandom responding to the word *good* on the first subliminal task.

Interestingly, removing the independent variable measuring differentiation of characters on the TAT card decreased the "explanatory power" of both the anger and splitting variables (result not shown). That is, removing TAT differentiation significantly lowered the incremental R^2 values associated with both anger and splitting score.

In trying to understand this phenomenon I noted that the coefficients associated with the TAT variable were negative, in contrast to the positive coefficients associated with anger and the splitting scale score. Furthermore, the TAT coefficients were negative before other variables were entered in a stepwise regression, and remained so throughout. Also, I noted that the explanatory power and statistical significance of the TAT variable increased as the anger variable and splitting scale score were entered in a stepwise manner.

This pattern of results would make sense if one assumed that nonrandom responding to the word *good* and the act of

significantly differentiating TAT characters both expressed a tendency to split, and if one further assumed that doing one lessened the need to do the other. That is, where splitting was expressed in a strongly nonrandom response pattern to the subliminally presented word *good* the psychological need to split the TAT characters diminished.

In other words, the more frequently a subject split, and the angrier she was, the greater the saliency of judgments about the "goodness" of the self. This manifested in an increase in nonrandom responding to the subliminally presented word *good*. However, an angry and frequent splitting subject who for some reason was unable to respond nonrandomly to the subliminally presented term (perhaps due to a very high perceptual threshold, or boredom with the task and resulting inattention) retained a tendency to split after the subliminal task, and was thus more likely to find expression in the TAT task.

Note that the temporal ordering of experimental tasks does not actually justify entering TAT as an independent variable in a test of improbability scores, because the latter data were derived first in time. Thus, to test my speculations appropriately I used multiple regression to test the effects of various variables, including the improbability weight of the response pattern to the word *good* on the first subliminal task, on differentiation of TAT characters.

All mood variables, splitting scale score, and measures of schematicism for the word *good* were also entered if *p* value associated with them was less than .2. Once all eligible items were entered, items were removed if the *p* value associated with them exceeded .25. This process resulted in the elimination of angry mood as an independent variable. However, as reflected in Table 18, splitting scale score was positively associated with TAT differentiation, as predicted, but nonrandomness of response to the subliminal word *good* was negatively associated with TAT differentiation, again as predicted. However, the incremental R^2 value associated with the improbability score for the word *good* is quite small (.03), lending little credence to my speculation.

Summary

In this chapter I described the procedures and hypotheses associated with Study 2, presented results from statistical tests of the hypotheses, and drew conclusions regarding the validity of the hypotheses.

Table 18. Results of regression of differentiation of TAT characters on schematic rating of the word *good*, improbability scores associated with responses to the word *good* on the 1st subliminal task, the mood variables, and splitting scale score, after removal of items with p values greater than 0.25.

Variable	*B*	*SE*	*Beta*	*t*	*p*
Schematicism	..... removed				
Schematicism^2	.0260	.0187	.2041	1.386	.1744†
Improbability Score for Word *good* on First Subliminal Task	-.1509	.0765	-.2937	-1.973	.0281‡
Boredom	5.4683	3.6960	.2135	1.480	.1477†
Splitting Scale Score	.2771	.1157	.3634	2.395	.011‡
Constant	9.1103	1.8252		4.992	.000‡

R =.290
Incremental R^2 due to splitting variable = .080
Incremental R^2 due to improbability variable =.031

† One-tailed *p*.
‡ Two-tailed *p*.

CHAPTER 4

CONCLUSIONS AND SUGGESTIONS FOR FURTHER RESEARCH

In this chapter I summarize the major findings, discuss important limitations of the study, and offer suggestions for further research in this area.

Summary of Results

The two experiments, taken together, provide fairly convincing evidence that college students obtaining low Gerson scale scores are likely to rate themselves more extremely on the global trait dimensions *good*, *bad*, *loving*, and *hateful* than comparable students obtaining high Gerson scale scores.

High Gerson scorers showed a greater tendency, in both studies, to respond inconsistently to trait terms presented perceptibly.

Gerson scale scores failed to covary significantly with reaction times for the words *good*, *bad*, or *hateful* on the tasks involving either perceptible stimuli or subliminal stimuli.

Interestingly, a greater tendency to split *was* associated with lower reaction time for the word *loving* on the first subliminal task.

[illegible] showed another information processing advantage for the words good and loving in the first additional task, being significantly more likely to respond consistently (i.e., nonrandomly) when repeatedly asked to guess whether the additional word described them or not.

Finally, nonrandomness of response to the word good in the first additional task was negatively associated with differentiation of TAT characters even as splitting scale scores remained positively associated.

Discussion

The results of the two studies do not generally support the idea that frequent splitters are more schematic on the traits good, bad, loving, and hateful. In fact, strictly speaking, the opposite appears to be true. Given the Fiske and Taylor (1984) criteria for assessing schematicism, frequent splitters were found to be significantly less schematic on the good/bad and loving/hateful dimensions than infrequent splitters.

Nevertheless, in a few instances frequent splitters appear to have enjoyed information processing advantages of a sort that would indicate greater schematicism. For example, they responded significantly less randomly to the words loving and good presented additionally than infrequent

splitters. On the other hand, the prediction that frequent splitters would respond more rapidly to stimuli, relative to reaction time for idiographic traits, was not supported.

Unfortunately, the design of the second study leaves room for interpreting the results in different ways. For example, subjects were asked to rate their own degree of schematicism on the trait dimensions only once, prior to other experimental procedures, including the mood inductions. Thus, one could speculate that frequent splitters failed to rate themselves more schematically because neither of their radically opposed self-schemata had been fully activated by environmental stimuli (e.g., a mood induction) when the measure was made.

If one imagines a borderline for whom neither the good self-schema nor bad self-schema has yet been activated, sense of self may be characterized by the sort of *identity diffusion* described by Kernberg (1975). Such a borderline could be conceived as existing in a temporary ego state marked by a very vague (but not complex) sense of self. In this *waiting state* an amorphous and impoverished self awaits an environmental stimulus to push it towards either the good pole or bad pole.

It is possible that the first (unhappy) mood induction played the role of environmental stimulus for many frequent splitters, pushing them out of the waiting condition to a more schematic condition. This, in turn, would have primed

them to respond nonrandomly to the key words *loving* and *good* (i.e. *schematically*), even though they rated themselves less schematically than infrequent splitters prior to the mood induction.

Such speculation exposes an important flaw in the design of the second study, namely the assumption that frequent splitters would be as stable in their subjective estimation of schematicism on important trait dimensions as infrequent splitters. The somewhat greater inconsistency of frequent splitters in responding *me* or *not me* to stimulus words presented perceptibly, as well as general clinical wisdom regarding frequent splitters, suggests that this assumption is unwarranted.

Follow-up research in this area could avoid relying upon the validity of such a dubious assumption by having subjects rate their schematicism on key traits several times during the course of a study. This would allow the stability of schematicism measures to be tested. If significant variability was found it could be factored into data analysis. For example, the researcher would have access to a measure of schematicism proximate to other measures under study.

The evidence for frequent splitters' increased nonrandomness of response to the subliminally presented words *good* and *loving* is impressive in my view. This evidence suggests that frequent splitters experienced more

priming on the good and loving dimensions, in turn suggesting a greater salience of these dimensions in the unconscious of frequent splitters. This conclusion would be stronger if nonrandom responding covaried with relatively faster reaction times, but this was true only in the case of the word *loving* on the first subliminal task.

I am inclined to give greater weight to the nonrandomness results given the problems of accuracy associated with the timing data (see Appendix A). Nevertheless, the inability of this study to demonstrate covariance of nonrandom responding and a reaction time advantage makes it impossible to conclude that frequent splitters acted more schematically than infrequent splitters once their schemata were activated by mood induction.

It could be argued, for example, that frequent splitters responded less randomly to subliminally presented trait stimuli precisely because they waited longer, on average, to respond. Length of exposure to the stimuli, rather than schematic priming, would then explain decreased randomness of response. Testing this quite plausible explanation of study results requires timing data more precise than those obtained in the present studies, as well as an *N* large enough to support ANOVA models with two or more factors.

Suggested Future Research

Some of the present results point to areas deserving of further study. Other results suggest areas which should be ignored in the next round of research in the interest of economy. Several suggestions for making future studies more powerful follow: first, timing data need to more precise. A device with dedicated timing switches should be used for recording responses, rather than the inherently less precise PC keyboard and software interrupt system used in the present research.

Second, schematicism on critical dimensions needs to be assessed more than once to permit testing of the hypothesis advanced above regarding movement of borderline splitters from an ego state involving aschematicism to one involving borderline self-schemata.

Third, a more powerful mood induction procedure is needed. The musical mood induction procedure described in Study 1 is one possibility. Another is the procedure described by Velten (1968), in which subjects are instructed to read a series of self-referential statements and asked to put themselves into the mood suggested by the statements.

So as not to confound the effects of mood and effects associated with temporal ordering of mood inductions, or simple fatigue, the ordering of mood inductions in a follow-up study should be randomly counterbalanced. Measures of

schematicism on significant traits need to be taken after each mood induction to ensure that any variability is noted and can be statistically controlled in subsequent data analysis.

Fifth, it seems advisable to replace free-form mood measurement with an instrument permitting subjects to be compared more directly, and quantitatively, on significant mood dimensions. My choice would be the *Visual Analogue Mood Scales* (VAMS), which requires subjects to mark positions on lines representing the degree to which they feel tired, anxious, despondent, sad, happy, angry, and apprehensive (McCormack, Horne, & Sheather, 1988). The scale is quickly administered, ensures comparable data between subjects--and within subjects over time if administered repeatedly.

Sixth, an increase in *N* is desirable so that procedure order can be counterbalanced to allow for more than two dichotomous ANOVA factors without producing empty cells.

Seventh, it would be useful to conduct follow-up studies with a well-defined borderline subsample serving as the frequent splitting group. This would help ensure maximal splitting and make the issue of generalizability of results more clearcut.

Finally, the combined length of study procedures needs to be reduced to avoid fatiguing subjects to the point at which results become suspect. Given the sparse results from the procedures involving perceptible presentation of

stimuli, they could be dropped in a future study. Their elimination alone would cut a 55 minute study to about 40 minutes. As much as another 10 minutes could probably be trimmed by eliminating deception and the props and procedures it required. Fortunately, deception would not be required to make the Velten or musical mood induction procedures work, another advantage in adopting them.

The standard trait terms *passive* and *assertive* could be eliminated from all tasks, as could the second set of idiographic trait terms. This would leave 4 standard trait terms and one set of idiographic terms, for a total of six. Each could then be repeated ten times during the subliminal tasks, rather than six, with no increase in the total time required for the procedure in Study 2. Repeating terms ten times would allow for more reliable estimates of reaction time and more powerful measurement of nonrandomness of response.

Procedures for the proposed follow-up study are summarized in Appendix E, which also lists hypotheses to be tested using the cleaner data this design should yield.

APPENDIX A

TIMING ERROR

Unable to procure the sophisticated timers usually used in research involving very rapid reaction times, I was compelled to use the PC as my timer. The computer program was written so that after initial presentation of a stimulus word a counting loop was initiated, continuing either until a time limit was reached (in which case a nonresponse was scored) or until an appropriate response key was depressed, interrupting and terminating the counting loop procedure.

There were several sources of measurement error. After timing dozens of count cycles from zero to 5,000,000, I found that the computer averaged 65,832 cycles per second, but calculations showed variability of as much as ±35 cycles per second. If this variance had been accrued evenly over the entire count cycle it would not have been problematic, inasmuch as an error of 35/65832 seconds (.00053 seconds) would be negligible. However, this error, attributable to unpredictable hardware-derived delay in command execution, was actually accrued only at the very beginning and end of measurement. Thus, error could amount to as much as 40 ms, a significant amount.[6]

Fortunately, this timing error was, for all practical purposes, random rather than systematic. This requires a bit of explanation. First, it is important to note that the error is associated with the point at which a timing interrupt was initiated in relation to where the CPU was in its processing or "clock" cycles. This means that the error, charted over time, would have a cyclical aspect to it. If the CPU was at clock step 1 the interrupt could not be enacted till after step 4, and the error would be greatest. If, however, the CPU was nearly completed with step 4, error would be minimal. Because the CPU cycled approximately 65,832 times per second (it took one full set of clock steps to complete one count cycle) this means that the amount of error which would accrue to a response varied from minimum to maximum 65,832 times per second. Because no person responds to experimental tasks of the sort in these experiments with timing precision approaching 1/65,832 seconds, the cyclical aspect of timing error here becomes an essentially moot point, and this source of error can be considered random.

However, there is some nonlinearity in the maximum ±40 ms contribution of timing error to observed reaction times, being larger in relation to the "true" reaction time when

---

[6] It should be understood that this degree of error represents a maximum, and that actual observed errors in test situations tended to be less than this.

the latter are very short than when they are very long (an additional reason to be suspicious of very short reaction times), but this nonlinearity is negligible above the 100 to 200 ms thresholds for reasonably plausible reaction latencies. Thus, even though the large contribution of timing error to general error variance in the reaction time studies is unfortunate in reducing the power of the inferential tests at my disposal, its essential randomness means it need not be taken into further account.

APPENDIX B

DETAILED PROCEDURES FOR STUDY TWO

We welcomed subjects to the experimental room and oriented them to the computer keyboard, particularly the five function keys (F1 through F5) marked in red, and the five function keys (F8 through F12) marked in green. These keys were arranged in a single horizontal line across the top of the keyboard.

Until this point the monitor screen was bright red with black lettering reading as follows in order to support the deception:

> AMERICAN PSYCHO-EDUCATIONAL LABORATORIES
> Testing Validation and Standardization Program
> for Test # 293.Sf5-17 is now ready to run.
> (copyright 1992)
> To be administered only at sites specifically
> authorized by APEL.

When we began the computer program a light blue screen with black lettering appeared. It read as follows:

> Thank you for agreeing to help us standardize a new form of cognitive testing. You and hundreds of other students are being tested now in study centers across the USA, using the new computerized methods you will see demonstrated today. (Press any GREEN key to continue)

The next screen read:

> Some of the procedures will appear very much like multiple choice testing with which you are already familiar. However, you will be asked to respond to these multiple choice questions as rapidly as possible. Speed of response will be measured and used, along with the correctness of your response, to evaluate how well you perform on each test you take today. (Press any GREEN key to continue)

The next screen read:

> Some of the tasks you will be asked to do today will seem quite different than anything you have ever done before. Once the testing is over the test administrator at your center will explain how each of these tasks is used to create a new index of cognitive functioning. (Press any GREEN key to continue)

The next screen read: "Occasionally the program will ask about your mood at that time. Knowing what your mood is will help us evaluate your performance more objectively. (Press any GREEN key to continue)".

The next screen read:

> Before we begin testing we want to know a little about you, and how you see yourself as a person, so we will be able to understand our results better. Some of this information will help us validate and standardize the tests. Some of this information will play a role in the tests, as you will later see. (Press any GREEN key to continue)

Next, the computer program prompted subjects to enter Year of birth, month of birth, day of birth, ethnic identification and year of college.

Next, the program prompted subjects to "please think of 3 words that best describe you mood at this moment" and asked them to enter the mood words by typing them on the keyboard.

After this a screen read as follows: "Please think of a quality that is strong in you and important to you. (Press any GREEN key to continue)".

The next screen read: "Now, referring to this quality that is strong in you and important to you please finish the following sentence with a single word: I am very _____." After subjects typed in their chosen adjective a follow-up screen read: "Please enter the word that is the opposite of the word x", where "x" represents the word subjects typed in a moment before. Once an antonym was typed in the cycle was repeated one more time, yielding four words provided by each subject.

These 4 adjectives were automatically combined in a temporary file with 6 standard adjectives used with all subjects: *good*, *bad*, *loving*, *hateful*, *assertive* and *passive*. Each adjective was then associated with a unique pseudo-random number generated by the program (the random number generator being seeded for each subject with an integer derived by algorithm from the number of seconds elapsed since midnight). This file was then sorted by the value of the pseudo random numbers, yielding a new file in which the adjectives were uniquely ordered for each subject.[7]

---

[7] Several other unique files of randomly ordered adjectives were created at the same time, for use in the various experimental procedures employing the adjectives, so that the ordering of adjectives varied from task to task, as well as from subject to subject. Some procedures (described below) required random presentation of each adjective 6

The experimenter now placed a special cover over the computer keyboard. This cover permitted access only to the red and green function keys described earlier. This cover, designed to limit subjects' response options to the function keys alone, would remain in place for all remaining procedures, with the exception of the mood checks, which were repeated several times and are described above.

The computer now prompted the subject to answer the following question: "How x are you?" (The x represents one of the randomly presented adjectives.) Five responses were possible, via the function keys: "Not x at all", "Not too x", "Average", Fairly x", and "Extremely x".

Immediately after a subject responded to this question the computer posed a follow-up question, its exact form dependent upon the subject's first response: "How important is it to you that you are ____ ?" (The blank represents one of the following phrases, in conformity with the subject's previous response: "Not x at all", "Not too x", "x to an average degree", "Fairly x", or "Extremely x".)

Possible responses included "Not at all important", "Not too important", "It is of average importance", "Fairly important", and "Extremely important". The point of these questions was to obtain a graded measure of the degree to which subjects viewed the adjectives self-schematically.

Once this procedure was completed a new screen appeared reading "APEL Reverse Digit Memory Test. (Press any GREEN key to continue)". The computer was programmed to beep once at this point in order to alert the researcher of the beginning of the "reverse digit test."

The next screen read:

> Please watch the screen carefully. A bracket will appear for a very brief time, followed by a number, which will also disappear after a few moments. Your job is to observe the number, keep it in mind for several seconds, and then indicate what the number is BACKWARDS. The first two trials are for practice. (Press any GREEN key to begin the test)

---

times, rather than a single time. The adjective files for these procedures were constructed in the following way: the original file of 10 adjectives was read, one adjective at a time, but each term was then written 6 times to another temporary file. The 60 terms on the temporary file (10 adjectives written 6 times) were each then associated with a psuedo-random number generated as described above. Thus, when this file was sorted by random number value the 60 terms were effectively randomized.

We had subjects repeat back instructions in their own words, as a way of insuring that there was no misunderstanding.[8]

Next, a pair of white brackets, appeared centered against an otherwise black screen. The brackets remained visible for approximately 500 ms. The screen was then blank for approximately 500 ms. A three digit number next appeared in the space which had been previously bracketed. The number remained visible for approximately 2000 ms. The screen was then blank for approximately 2000 ms. Then five three-digit numbers appeared at the center of the screen, each one preceded by the name of a function key (F1 through F5).

One, and only one, of the numbers represented the original number with the digits in reverse order. We checked, during the first two practice trials using three-digit numbers, to make certain subjects understood how to judge the correct answer and how to associate their answer with the appropriate function key.

After the "practice trials" ten additional target numbers were presented, ranging upward in size from three digits to nine digits. During pretesting nine digits had proved sufficient to insure failure at the task. This was important in making the negative feedback subjects received after the reverse digit task credible.

After the last reverse digit trial a screen appeared for approximately 5 seconds reading "Please wait while the computer calculates your performance in relation to other students of your age."

The next screen appeared automatically, with text at the top indicating that "this graph shows your combined performance (accuracy plus speed) compared to other students your age." Below the text a line stretched horizontally across the screen, with the word "poor" labelling its left end, the word "excellent" labelling its right end, and the word "average" labelling the middle. A vertical line, perpendicular to the horizontal one, and touching the latter to the left of the word "average", roughly 1/3 the way to the word "poor", was labelled "Your performance fell about here. (Below Average)". This screen could not be erased until approximately 7 seconds had elapsed, after which additional words appeared at the bottom of the screen: "(Press any GREEN key to continue)".

The next procedure, the so-called "intuition test", was introduced this way by the computer:

---

[8] This was important because we wanted subjects to feel they had done their best on this task, in order to enhance the effect of the negative feedback all subjects received about their performance at the end of the task.

Now you will take a test of INTUITION, which is important in many types of problem solving. In a moment you will see flashes of light on the screen. These flashes hide words that you probably won't be able to read at a conscious level. You WILL be able to register the words unconsciously, however. Then you WILL be able to make guesses about the word you unconsciously registered by using your intuition. (Press any GREEN key to continue)

The next screen read:

A moment before the words are flashed you will see a set of brackets on the screen. This is where the words will be flashed. Please watch the brackets and flashes carefully. Then, using your intuition, guess whether the hidden word describes you as a person. Do not hesitate! Make your guess as rapidly as possible! If you guess the hidden word DOES describe you press a GREEN key. If you guess the hidden word DOES NOT describe you press a RED key. (Press any GREEN key to continue)

An audible tone alerted researchers to check with subjects at this point, to insure that the latter understood the instructions. Based on the questions subjects asked here, it was often necessary for us to emphasize that subjects were unlikely to read the words that would be flashed, and would instead have to use their "gut feeling" to correctly guess whether the hidden word described them.

Next, the computer presented white brackets on the black monitor screen for approximately 2000 ms, followed by about 500 ms of blank screen. Next, to serve as frontmasking, a sequence of twelve "greater-than signs" (>>>>>>>>>>>>), presented for 50 ms, alternated with an equally long sequence of "less-than signs" (<<<<<<<<<<<<), flashed for the same length of time, over a total period of 1000 ms. Then the target word, one of ten stimulus adjectives described above, was printed on the screen for 350 ms. Backmasking followed, consisting of the alternating greater-than sign and less-than signs for another 1000 ms.[9]

---

[9] Backmasking was necessary because the phosphors comprising pixels on a monitor continue to glow for some period after they are "commanded" to turn black, resulting in an afterimage of unacceptable duration. Backmasking obliterates this afterimage, making the timing of presentation reasonably precise. In extensive experimentation prior to the study I determined that frontmasking helped "subjects" focus on the correct area of the screen, and reduced their conscious recognition of

Timing started at the presentation of the target word, and ended with either a red keypress or green keypress. A keypress blackened the screen for 3000 ms, at which point the next trial began. The computer automatically recorded the nature of the response and reaction time.

At the end of the intuition procedure subjects were once again asked to provide three words describing their present mood. It was explained that they could use the same mood adjectives they used before, if they wished, indicating any change of degree with the symbols + or -.

Next, a tone sounded and the screen went completely blank. We intervened, informing subjects that they should turn their chair around to face the experimenter. Subjects were told that they would next be shown a picture and asked to construct a story about it:

> Now I will show you a picture. First, I would like you to give a name to each of the persons in the picture. Next, I would like you to tell me a story about what is happening in the picture, beginning with what happened before the picture, then telling me what is happening now, and finally what you think will happen in the future, after the scene shown in the picture.

We showed subjects card 9GF from the Thematic Apperception Test, which Groth-Marnat (1990, p.338) described this way: "a woman in the foreground is standing behind a tree. Below her is another woman running along a beach."

I intended the TAT task to provide an independent measure of splitting with which to compare data from the intuition procedure. Even though normally it would be desirable to use several TAT cards in order to detect splitting, or any other defense, time did not allow the administration of multiple cards. I chose card 9GF because it depicts two characters of very similar appearance, thus providing suitable grist for a split projection, without pulling for it too directly.

I did not score the stories. Instead, I had subjects return to the computer, which first prompted them to type in a name for each TAT character, and then asked questions of the form "How x is A?" and "How x is B?" (where *x* represents one of the ten stimulus adjectives and *A* and *B* represent the respective TAT character names).

---

stimuli without affecting their ability to "guess" correctly whether they had previously identified the word as *me* or *not me*. Length of stimulus presentation in this study was also based on experimentation with friends and colleagues prior to the study.

Response choices were: "Not x at all", "Not too x", "Average", "Fairly x", and "extremely x". The computer assigned these responses numerical values from 0 to 4, and computed the difference between A and B's scores for each of the ten stimulus adjectives. The degree to which the character scores differed would later be interpreted as evidence of splitting projected onto the TAT card.

Next, the computer presented this screen text:

> In order to get to know your innate response speed we will now ask some easy questions about you as a person. In a moment you will see descriptive words on the screen and your task is to decide, as rapidly as possible, whether the term describes you or does not describe you... Please respond as rapidly as possible by pressing a key outlined in GREEN if the word does describe you as a person or a key outlined in RED if the word does not describe you. Remember: if the word DOES describe you press a GREEN key. If the word DOES NOT describe you press a RED key. (Press any GREEN key to continue)

The next procedure was similar to the intuition task except that the 60 presentations of the stimulus adjectives were fully perceptible and no front or back masking was used. Each term was presented in letters 1.3 cm high for as long as 5000 ms, or until subjects pressed a red or green key, terminating the trial. The words were presented in 3 blocks of 20 trials, with 20 second rest periods between blocks. Again, the computer recorded responses and response latencies.

Next, the computer presented a screen reading: "APEL LATERAL LOCALIZATION TEST PROTOTYPE". Subjects were instructed to press one of four function keys (F1 through F4) whenever a small cross symbol was flashed on the screen, F1 if the cross was in the leftmost quadrant of the screen, F2 if it was in the next quadrant to the right, and so on. The computer instructed subjects not to press a key whenever distractor symbols were flashed.

Symbols were next presented for 1000 ms at random locations along the horizontal line passing through the screen's center. The researcher observed the first few key presses to insure that subjects understood the instructions. After 100 symbols (including 20 crosses) were presented the screen blanked for a few seconds. Then a graph appeared, identical to the one used in the reverse digit task, with the exception that in this case it indicated that subjects' performance was "very good".

The computer again queried for mood, repeated the subliminal "intuition test" and the task involving perceptible stimulus adjective presentation, and queried mood a final time.

The computer program terminated with a tone and a blank screen. We congratulated subjects for finishing the testing. Still in role, we asked for feedback regarding the test procedures. The chief purpose of this query was to give subjects who saw through the experimental deception a good opportunity to say so, prior to this being directly revealed to them. No subject expressed doubt about the cover story at this point. We asked which procedures subjects liked most and least. Subjects typically reported liking and/or doing well on the cross symbol test, and not liking and/or not doing well on the reverse digit test. This provided additional evidence that subjects did not see through the deception.

Finally, we asked what subjects thought of the subliminal task, and most subjects reported having detected only a few, if any words. Those subjects who reported having detected words almost invariably reported that the word they detected was the first schematic adjective they themselves provided. In only two instances did subjects report seeing a standard trait term, and then only once or twice during the entire procedure. Thus, adjective presentation seems to have been truly subliminal as intended.

We then informed subjects of the deception and fully debriefed them. We took special care to make certain they realized that their performance had never been graded in relation to others. We then dismissed subjects.

APPENDIX C

CATEGORIZATION OF MOOD TERMS IN STUDY TWO

Subjects in Study 2 were coded 0 or 1 on mood variables representing anger, sadness, boredom, and anxiety. Zero was the default value on these variables. Subjects were also coded 0 if their mood responses gave positive indication of *not* being in the respective mood (e.g., the response *happy* resulted in the sadness variable being coded 0).

Mood responses that resulted in a coding of 1 for angry mood included *angry*, *annoyed*, *irritated*, and *mad*.

Mood responses that resulted in a coding of 1 for sad mood included *depressed*, *somber*, *pessimistic*, and *blue*.

Mood responses that resulted in a coding of 1 for bored mood included *bored*, *restless*, *indifferent*, *tired*, and *unenthusiastic*.

Mood responses that resulted in a coding of 1 for anxious mood included *nervous*, *geared up*, *apprehensive*, *anxious*, *hyper*, *timid*, *stressed*, and *worried*.

APPENDIX D

THE IMPROBABILITY SCORE FROM STUDY TWO

Imagine that a subject responds completely randomly to all six subliminal presentations of a given word. "Completely randomly" means that there is no priming effect in operation, so that the "guess" really is a random guess. Furthermore, complete randomness dictates an absence of any response bias. Thus, for example, the subject cannot have an inherent bias toward responding *me* or *not me*.

Under these conditions the expected results of randomized responding would be equivalent to the expected results if we flipped a fair coin for each response, heads perhaps representing *me* and tails representing *not me*. The binomial theorem is easy to apply in such a case. Table 19 presents the probability of achieving so many *me* responses and so many *not me* responses, in any order, out of six trials, the probabilities determined by the formula provided in Blalock (1979),

$$\frac{N!}{r!(N-r)!} * p^{r} * q^{N-r},$$

where N equals the total number of trials (6 in this case), r equals the number of *me* responses, p refers to the odds of getting a *me* response (1/2 in this case) and q equals the odds of getting a *not me* response (again 1/2 in this case).

Table 19. Binomial probabilities associated with achieving different numbers of me and not me responses, in any order, out of 6 trials, assuming completely random responding.

Number of *me* Responses	Probability
0	1/64
1	1/10.666...
2	1/4.266...
3	1/3.2
4	1/4.266...
5	1/10.666...
6	1/64

In this unrealistically simple case note that the probabilities are symmetrical around the most highly expected result of three *me* responses out of six trials. Testing a null hypothesis of random responding with one subject is easy in this case, because the probability of an observed response pattern if the null hypothesis is true equals the probabilities at the right side of Table 19. Thus, the only patterns which would permit us to reject the null hypothesis of randomness would be zero *me* responses or six *me* responses, because only these patterns are associated with *p* less than .05 (*p*=.016, to be more precise).

In the present study the problem was not to make inferences about individual subjects, but rather to see if frequent splitters responded less randomly to the words *good* and *bad* than infrequent splitters. To do so, I took the reciprocal of the *p* values associated with a subject's observed response pattern, yielding, in essence a weight corresponding to the improbability of that subject's response pattern assuming that the responding is completely random. Averaging these weights among frequent splitters and infrequent splitters, I obtained a measure of the average improbability of the response patterns.

Computation of improbability scores for individuals actually required correction for response bias, because any tendency to respond *me* more than *not me*, or vice versa, would skew *p* values.

Although we have no absolute measure of response bias, we do have a handy value with which to estimate it, namely the proportion of *me* to total responses, over all presentations of all trait adjectives in a given task. The logic here is that because trait adjectives were paired in opposites (e.g., *good* and *bad*, *loving* and *hateful*, etc.) the disavowal and endorsement of adjectives should balance if there is no response bias. Under this assumption, anything other than a .5 proportion of *me* to total responses reflects response bias. The observed proportion of *me* responses was entered in the binomial equation as p, and the observed proportion of *not me* responses was entered as q.

APPENDIX E

PROCEDURE LIST AND HYPOTHESES FOR PROPOSED FOLLOW-UP STUDY

Procedures:

1) Introduction and informed consent
2) Visual Analogue Mood Scale (VAMS)
3) Subjects provide idiographic schematic trait terms
4) Rating of self-schematicism (RSS) of trait terms
5) Happy or sad mood induction (by constrained random assignment)
6) VAMS
7) RSS
8) Subliminal *me/not me* judgment task
9) TAT task
10) Happy or sad mood induction (opposite of #4)
11) VAMS
12) RSS
13) Subliminal *me/not me* judgment task
14) TAT task
15) Debriefing

Major Hypotheses:

1) Borderline subjects (BS) will show more variability in VAMS.
2) BS will show more variability in RSS.
3) For BS only, variability in VAMS will be positively associated with variability in RSS.
4) BS will achieve higher improbability scores for words *good*, *bad*, *loving*, and *hateful* on subliminal tasks.
5) Within subjects, improbability scores for different trait terms will be negatively associated with reaction time.

REFERENCES

Akhtar, S. and Byrne, J. P. (1983). The concept of splitting and its clinical relevance. *The American Journal of Psychiatry, 140,* 1013-1016.

Armstrong, J. G., & Loewenstein, R. J. (1990). Characteristics of patients with multiple personality and dissociative disorders on psychological testing. *The Journal of Nervous and Mental Disease, 178,* 448-454.

Beck, A. T. (1979). *Cognitive therapy and the emotional disorders*. New York: International Universities Press.

Berman, E. (1981). Multiple personality: Psychoanalytic perspectives. *International Journal of Psycho-analysis, 62,* 283-300.

Blalock, H. M. (1979). *Social statistics*. New York: McGraw-Hill.

Bion, W. R. (1959). *Experience in groups*. New York: Basic Books.

Blanck, G. and Blanck, R. (1974). *Ego psychology: Theory and practice*. New York: Columbia University Press.

Breuer, J. & Freud, S. (1955). Studies on hysteria. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 2. pp. 1-305). London: Hogarth Press. (Original work published 1893-1895)

Cashdan, S. (1988). *Object relations therapy*. New York: W. W. Norton.

Clark, D. M. (1983). On the induction of depressed mood in the laboratory: Evaluation and comparison of the Velten and musical procedures. *Advances in Behavior Research and Therapy, 5,* 27-49.

Collins, R. (1983). *Rorschach correlates of borderline personality*. Unpublished doctoral dissertation, University of Toronto.

Cooper, S. H. and Arnow, D. (1985). An object relations view of the borderline defenses: A Rorschach analysis. In M. Kissen (Ed.), *Assessing object relations* (pp. 143-171). New York: International University Press.

Cooper, S. H., Perry, J. C., and Arnow, D. (1988). An empirical approach to the study of defense mechanisms: I. reliability and preliminary validity of the rorschach defense scales. *Journal of Personality Assessment, 52,* 187-203.

Delibes, L. (Composer), Orchestre de la Suisse Romande (Orchestra), & Bonynge, R. (Conductor). (1989). Mazurka. *Coppélia* [Ballet on compact disc]. London: The Decca Record Company.

Dervin, D. (1991). Splitting and fragmentation of group-fantasy during the Reagan years. *The Journal of Psychohistory, 18,* 283-292.

Fiske, S. T. (1982). Schema-triggered affect: Applications to social perception. In: M. S. Clark and S. T. Fiske (Eds.) *Affect and cognition: The seventeenth annual Carnegie symposium*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. New York: McGraw-Hill.

Freud, A. (1948). *The ego and the mechanisms of defence*. London: The Hogarth Press.

Freud, S. (1955a). The defence neuro-psychoses. In J. Strachey (Ed. & Trans.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (Vol. 3, pp. 45-61). London: The Hogarth Press. (Original work published 1894)

Freud, S. (1955b). Outline of psycho-analysis. In J. Strachey (Ed. & Trans.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, (Vol. 23, pp. 144-207). London: The Hogarth Press. (Original work published 1940)

Freud, S. (1955c). Splitting of the ego in the process of defence. In J. Strachey (Ed. & Trans.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud*, (Vol. 23, pp. 275-278). London: The Hogarth Press. (Original work published 1940)

Gabbard, G. O. (1989). Splitting in hospital treatment. *American Journal of Psychiatry, 146*, 444-451.

Gergely, G. (1992). Developmental reconstructions: Infancy from the point of view of psychoanalysis and developmental psychology. *Psychoanalysis and Contemporary Thought, 15*, 3-55.

Gerson, M. J. (1984). Splitting: The development of a measure. *Journal of Clinical Psychology, 40*, 157-162.

Glassman, M. B. (1986). Splitting: Further contributions to the development of a measure. *Journal of Clinical Psychology, 42*, 895-904.

Gouaux, C. (1991). Unjust wars as borderline phenomena. *Journal of Psychohistory, 19*, 109-113.

Gough, H. G. (1952). *The adjective check list*. Palo Alto, California: Consulting Psychologists Press.

Greene, L. R., Rosenkrantz, J., and Muth, M. Y. (1985). Splitting dynamics, self-representations and boundary phenomena in the group psychotherapy of borderline personality disorders. *Psychiatry, 48*, 234-245.

Greene, L. R., Rosenkrantz, J., and Muth, M. Y. (1986). Borderline defenses and countertransference: Research findings and implications. *Psychiatry, 49*, 253-264.

Grotstein, J. S. (1981). *Splitting and projective identification*. New York: Jason Aronson.

Hartmann, H. (1958). *Ego psychology and the problem of adaptation*. New York: International Universities Press.

Horowitz, M. J. (1977). Cognitive and interactive aspects of splitting. *American Journal of Psychiatry, 135*, 549-553.

Jones, E. (1950). Introduction. In E. Jones (Ed.), *Contributions to psycho-analysis, 1921-1945: Melanie Klein* (pp. 9-12). London: Hogarth Press.

Kenealy, P. (1988). Validation of a music mood induction procedure: Some preliminary findings. *Cognition and emotion, 2*, 41-48.

Kernberg, O. (1975). *Borderline conditions and pathological narcissism*. New York: Jason Aronson.

Kernberg, O. (1984). *Severe personality disorders; Psychotherapeutic strategies*. New Haven: Yale University Press.

Kernberg, O. (1988, July). *Borderline and narcissistic patients*. Lecture at the Albert Einstein College of Medicine, Cape Cod Institute, Yarmouth, MA.

Klein, M. (1950). The development of a child. In E. Jones (Ed.), *Contributions to psycho-analysis, 1921-1945: Melanie Klein (pp. 13-67)*. London: Hogarth Press.

Klein, M. (1955). On identification. In M. Klein, P. Heimann, & R. E. Money-Kyrle (Eds.), *New directions in psycho-analysis* (pp. 141-175). London: Tavistock Publications.

Kohut, H. (1971). *The analysis of the self*. New York: International University Press.

Kuiper, N. A. (1981). Convergent evidence for the self as a prototype: The 'inverted-U RT effect' for self and other judgments. *Personality and social psychology bulletin, 7*, 438-443.

Kuiper, N. A., & Rogers, T. B. (1979). Encoding of personal information: Self-other differences. *Journal of personality and social psychology, 37*, 499-514.

Kuiper, N. A., Olinger, L. J., MacDonald, M. R., & Shaw, B. F. (1985). Self schema processing of depressed and nondepressed content: The effect of vulnerability to depression. *Social Cognition, 3*, 77-93.

Ledbetter, S. (1987). Liner notes. In Prokofiev, S. (Composer), Los Angeles Philharmonic Orchestra (Orchestra), & Previn, A. (Conductor), *Alexander Nevsky; Lieutenant Kije* [Film score on compact disc]. Cleveland, OH: Telarc International Corporation.

Lerner, P. M. (1991). *Psychoanalytic theory and the Rorschach*. Hillsdale, NJ: The Analytic Press.

Lerner, H. D., and Lerner, P. M. (1982). A comparative study of defensive structure in neurotic, borderline, and schizophrenic patients. *Psychoanalysis and contemporary thought, 5*, 77-115.

Linehan, M. M. (1987). Dialectical behavior therapy for borderline personality disorder: Theory and method. *Bulletin of the Menninger clinic, 51*, 261-276.

Linehan, M. M. (1993). *Cognitive-behavioral treatment of borderline personality disorder*. New York: The Guilford Press.
Linville, P. W. (1982). The complexity-extremity effect and age-based stereotyping. *Journal of personality and social psychology, 42*, 193-211.
Linville, P. W. (1985). Self complexity and affective extremity: Don't put all your eggs in one basket. *Social cognition, 3*, 94-120.
Linville, P. W. (1987). Self-complexity as a buffer against stress-related illness and depression. *Journal of personality and social psychology, 52*, 663-676.
Mahler, M. (1971). A study of the separation-individuation process: And its possible application to borderline phenomena in the psychoanalytic situation. *The psychoanalytic study of the Child, 26*, 403-424.
Mahler, M., Pine, F. and Bergman, A. (1975). *The psychological birth of the human infant*. New York: Basic Books.
Markus, H. (1977). Self-schemata and processing information about the self. *Journal of personality and social psychology, 35*, 63-78.
Markus, H. and Nurius, P. (1986). Possible selves. *American psychologist, 41*, 954-969.
Masterson, J. (1981). *The narcissistic and borderline disorders*. New York: Brunner/Mazel.
McCormack, H. M., Horne, D. J., & Sheather, S. (1988). Clinical applications of visual analogue scales: A critical review. *Psychological medicine. 18*, 1007-1019.
Millon, T. (1987). *Millon clinical multiaxial inventory (2nd ed.)* Minneapolis, MN: Computer Systems.
Morrison, T. L., Greene, L. R., & Tischler, N. G. (1985). Manifestations of splitting in the large group. *The journal of social psychology, 125*, 601-611.
Myers, J. L. & Well, A. D. (1991). *Research design and statistical analysis*. New York: Harper Collins.
Ogden, T. (1991). *Projective identification and psychotherapeutic techniques*. Northvale, NJ: Jason Aronson.
Parrott, W. G. (1991). Mood induction and instructions to sustain moods: A test of the subject compliance hypothesis of mood congruent memory. *Cognition and emotion. 5*, 41-52.
Pretzer, J. (1990). Borderline personality disorder. In A. T. Beck and A. Freeman (Eds.) *Cognitive therapy of personality disorders*. New York: The Guilford Press.
Prokofiev, S. (Composer), Los Angeles Philharmonic Orchestra (Orchestra), & Previn, A. (Conductor). (1987). Russia, Beneath the Yoke of the Monguls. *Alexander Nevsky; Lieutenant Kije* [Film score on compact disc]. Cleveland, OH: Telarc International Corporation.

Rappaport, D., Gill, M. M., and Schafer, R. (1968). *Diagnostic psychological testing*. New York: International University Press.

Ross, J. M. and Dunn, P. B. (1980). Notes on the genesis of pathological splitting. *International Journal of Psycho-Analysis, 61*, 335-349.

Searles, H. F. (1986). The countertransference with the borderline patient. In Stone, M. H. (Ed.), *Essential papers on borderline disorders; One hundred years at the border* (pp. 498-526). New York: New York University Press.

Spielberger, C.D., Gorsuch, R.L., & Lushene, R.E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Counseling Psychologist Press.

Velten, E. (1968). A laboratory task for induction of mood states. *Behavior research and therapy. 6*, 473-482.

Van-Der Keshet (1988). *Anorexic patients and ballet students: A Rorschach analysis*. Unpublished doctoral dissertation. University of Toronto, Toronto.

Volkan, V. (1987). *Six steps in the treatment of borderline personality organization*. Northvale, NJ: Jason Aronson.