

DELROY L. PAULHUS

University of British Columbia
Vancouver

THREE EXPERIMENTAL PARADIGMS FOR EXAMINING DEFENSE MECHANISMS

I describe three experimental paradigms that I have developed to investigate defense mechanisms. All three exploit techniques borrowed from recent developments in cognitive psychology. The first paradigm involves manipulating emotional arousal to examine its effects on self-descriptions: In general, arousal increases the positivity of current self-evaluation. The second paradigm involves the use of threatening distractor-words to influence subjects' self-reports: Again, increases in self-positivity are typically observed. The third paradigm involves biasing subjects' self-reported traits by telling them that certain traits are diagnostic of some future positive event (e.g., happiness in middle age) or some negative event (e.g., failed marriages). Although their trait reports can have no effect on the outcome, subjects tend to claim traits associated with positive future outcomes and disavow those associated with negative outcomes. Future applications are discussed.

This article was developed from two papers read at the Gulf-Bridging Symposium on Defense Mechanisms, Self-Deception, and Cognitive Error held in Nieborów, Poland, 1988. One recurring theme at the symposium was the testability of theories of psychological defense. In the past, this provocative concept has often been dismissed by experimental psychologists as being beyond the pale of scientific analysis. That view was not shared by the participants at the symposium. My contribution to the symposium and to the present issue is a description of my own attempts at experimentally anchoring the often elusive constructs involved in psychological defense.

All three methodologies to be discussed share a number of features. For example, all three involve a demonstration of affective bias in self-reports. They all permit a controlled manipulation of individuals' self-descriptions by creating some emotional state that motivates a distorted self-presentation. The metho-

Please send requests for reprints to Delroy L. Paulhus, Department of Psychology, University of British Columbia, Vancouver, Canada V6T 1W5.

dologies are all couched in the language of modern information processing theories. Finally, all three are flexible enough to permit the testing of a variety of issues.

PARADIGM I: EMOTIONAL AROUSAL AND SELF-DESCRIPTIONS

In a recent review paper, I developed a theory of dynamic complexity to explain certain effects of emotional arousal on self-descriptions (Paulhus & Lim, 1987). The theory was supported by a program of 13 experiments. In brief, the theory proposes that arousal reduces one's cognitive complexity because of reduced capacity for multidimensional analysis. Because, in social domains, evaluation is invariably the most important dimension of analysis, it is the last dimension to be discarded under arousal. The result is that evaluation appears to polarize: positively-valued targets become more positive and disliked targets become more negative. In general, then, emotional arousal tends to exaggerate evaluations of social targets.

The theory applies equally to ratings of the self and others. Thus, the effect of emotional arousal tends to be an exaggeration of one's typical self-evaluation (i.e., self-esteem). Since most individuals have a generally favorable self-esteem, their self-evaluation tends to increase.

A variety of arousal agents and dependent variables have been used to confirm these propositions. One experiment examined the effects of an impending exam on students self-evaluations (Paulhus & Lim, 1987, Study 1). Students were asked to rate themselves on a series of trait adjectives designed to tap evaluation. They were tested three times on parallel forms of this self-evaluation measure: a week before the class exam, 15 minutes prior to the exam, and two-weeks later.

The results showed an increase in self-evaluation just prior to the exam. Scores before and after were significantly lower but not significantly different from each other.

In another experiment, we used white noise as an arousal agent. This manipulation is often used to create a relatively neutral form of arousal (e.g., Hamilton, Hockey, & Rejman, 1977). The dependent variable was the strength of the evaluation dimension inferred from similarity ratings of self-roles. Multidimensional scaling was used to verify that evaluation was the primary dimension in these ratings and to provide an index of its importance. The results showed that the importance of evaluation in self-ratings was amplified by the white noise. Individuals rated their liked roles higher and their disliked roles lower.

Subsequent studies with white noise showed that direct measures of self-evaluation, as used in the first study described above, were also responsive to the noise manipulation. Specifically, self-evaluation scores increased under noise.

Other studies compared the effect of these manipulations on high- and low-self-esteem individuals. This required the pre-selection of enough low self-esteem individuals to permit statistical analysis. As predicted by the theory,

the polar
showed i
decrease
current f

Much
effects. T
Nonethe
was the f
most obs
my colle
than aut

We
complex
point wa
of others
of more
we conc
self-este
exhibitin

The
compon
use of p
Second,

The
middle c
me" to e
as the tr
later". T
innocuo

Two
paralle
to incre
Moreov
whereas

Ano
distract
reached
subjects

After

the polarization pattern was evident. Although high self-esteem individuals showed increased self-evaluation, low self-esteem individuals actually showed decreased self-evaluation. It is not yet clear whether arousal affects subjects' current feelings or only their reports thereof.

Much effort was expended to determine the psychological mediator of these effects. The theoretical predictions were based on known effects of arousal agents. Nonetheless, the demonstration of arousal is a difficult task. Most problematic was the fact that there is still little consensus on the nature of arousal, although most observers concede that it is multidimensional. After a series of experiments my colleagues and I concluded that the effective mediator was cortical rather than autonomic arousal (Paulhus & Lim, 1987).

We drew a number of conclusions about the application of dynamic complexity theory to psychological defense (Paulhus & Suedfeld, 1988). One point was that the representation of the self behaved much like the representation of others under arousal. Thus defense mechanisms should be considered a subset of more general theories about how emotion affects cognitive processing. Second, we concluded that the defense system is ineffective in individuals with low self-esteem. These people are actually worse off under emotional conditions, exhibiting much negative affect toward themselves as well as others.

PARADIGM II: THE DISTRACTOR PARADIGM

The second experimental paradigm I wish to describe has several face-valid components of defense mechanisms. First, threat is directly manipulated by the use of psychologically threatening distractor words on a computer screen. Second, the dependent variable is self-evaluation, which is central to defense.

The experimental procedure is as follows. Trait words are presented in the middle of a micro-computer screen. Subjects are asked to respond "me" or "not me" to each of the traits. Off to the side a distractor word appears at the same time as the trait word. Subjects are told to ignore the word "We will get back to that later". The distractor words are either threatening (e.g., *penis, kill, blood*) or innocuous (*sofa, justice*).

Two such studies were reported in Paulhus and Levitt (1987). The effects paralleled those claimed for defense mechanisms. Threatening distractors tended to increase the claiming of positive traits and the denial of negative traits. Moreover, response latencies to positive or negative traits decreased under threat whereas latencies to neutral traits increased under threat.

Another clearcut finding was that subjects were vigilant for threatening distractors, contrary to a perceptual defense hypothesis. This conclusion was reached on the basis of signal detection analyses of a recognition test given to subjects after the experimental procedure.

After ruling out a number of alternatives, Paulhus and Levitt (1987) offered

two possible mechanisms for such effects. One was fast-rising arousal. Traditional Hull-Spence learning theory held that arousal increased dominant responses and decreased subordinate responses. Given that desirable responses tend to be dominant, the increased claiming of positive and denial of negative traits may be simply a Hull-Spence effect.

Alternatively, we suggested that a pure attentional theory could explain the results. Given that threatening words attract more attention, it might be that subjects made a more superficial analysis before responding. Therefore we conducted a parallel study using a non-emotional distractor, namely, counting digits (Paulhus, Graf, & Van Selest, 1989). When the digit task was made very difficult, the results paralleled those obtained with emotional distractors — that is, increased claiming of desirable traits and decreased claiming of undesirable traits. Hence, diverting subjects' attention during self-descriptions increases dominant or autonomic responding. Alternatively, under distraction, subject might consider only the desirability of the trait, an analysis which is known to be quicker than trait ratings (see Ferguson, Rule & Carlson, 1983).

In another study (Paulhus & Murphy, 1987), the affective distractors caused a *decrease* in positivity in subjects instructed to "fake good" and an *increase* in positivity in subjects instructed to "fake bad". In both cases subjects moved to a moderately-positive level of self-presentation when the distractors were affective. To explain all of these findings, a more general theory was required.

A theory of autonomic and controlled self-presentation was recently laid out in detail (Paulhus & Baumeister, 1988). The model exploits the distinction made by cognitive psychologists between autonomic and controlled processes (Posner & Snyder, 1975; Schneider & Shiffrin, 1977). Autonomic processes are those that are so well-practiced that they operate without attention. They are effortless and cannot be stopped once they are under way. For example, the semantic processing (reading) of a word is autonomic. In contrast, controlled processes require attention to proceed. For example, counting backwards by threes or recounting one's childhood are controlled processes because they stop when attention is withdrawn.

To the extent that the attention required for a controlled task is drawn away (for example, by an affective event), then that task will continue in autonomic mode. As Simon (1986) has argued, the preeminent switching device is affect. It tells you which task is most important and therefore deserves your attention.

The automatic/controlled distinction has already been applied to social phenomena (for a review, see Bargh, 1984). I propose that self-presentation involves a system of such automatic and controlled processes. Self-presentation is defined here as the tendency to describe oneself in favorable terms. At times, self-presentation is controlled: that is, it involves purposeful and tailored self-descriptions. One considers the perspective of the current audience and the particular strategies that would best create a desirable impression. Such self-presentation requires maximal attention.

At other times, people fire off positive self-descriptions without thought or memory search. This autonomic self-presentation is likely to occur in "mindless" self-descriptions or under a time shortage or information overload. The self-descriptions that appear here are highly practiced from a lifetime of repetition. Therefore they can be emitted quickly with minimal attention. We are encouraged from childhood to say positive things about ourselves and continue to do so as adults. Hence, this autonomic self-report is generally positive and relatively consistent over time.

Affective stimuli can switch self-presentation from the controlled mode into the autonomic mode. Social threats, for example, make some people brag and other people derogate themselves. I suggest that both types are emitting the autonomic self, which in some people is positive and in other people, negative. The common cause is a loss of attention to appropriate self-presentation caused by emotional arousal. Under negative emotions, people may be devoting their attention to affect management (Clark & Isen, 1982).

Self-presentation is often required concurrently with other objectives. Indeed, there are often multiple objectives involved in the self-presentation itself. Therefore, attention must continually be switched from one objective to another. To the extent that attention is available for a self-presentation, it will be controlled. If attention is drawn elsewhere but the self-presentation continues, it will move to the automatic level. These arguments suggest the model presented in Figure 1. The automatic level of self-presentation represents a default level that is reverted to when any form of controlled self-presentation is disrupted. In individuals with high self-esteem, the automatic level is positive, although not as positive as the perfect self-presentation possible in faking. The honest self-report is also a form of controlled processing because it requires memory search, reflection, and evaluation. Because honest reflection often yields negative information, this level will be lower than the automatic level.

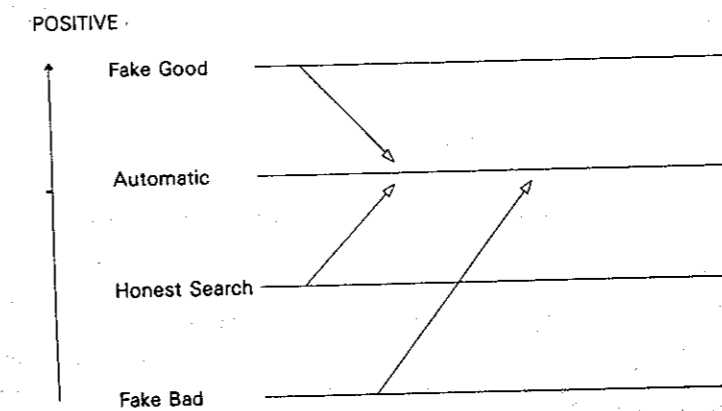


Fig. 1 Modes of Self-Presentation

This research program continues on two fronts. In the first subprogram, some basic issues about the nature of the affective distraction paradigm are being tackled: Is the loss of attention an overt or covert phenomenon? What is automatic — the practiced self or simply desirable responses?

PARADIGM III: AN INDEX OF MOTIVATIONAL BIAS

The third experimental paradigm was developed from a procedure used by Quattrone and Tversky (1983). They induced a motivational bias in subjects by associating a negative outcome with a particular self-report. Subjects were given the cold-processor test and asked to report the level of pain they experienced. Some subjects were told beforehand that people who experience pain in that situation tend to have constitutional factors predisposing them to heart attacks and other physical disabilities. These subjects reported less pain and kept their arms in the cold water longer.

In collaboration with Gordon Murphy, I developed a more versatile index titled the "Index of Motivational Bias". Subjects rate themselves on 40 traits including the Big Five dimensions of personality (Norman, 1963). Any motivational instructions may be used to bias subjects' responses. In our standardized version, we use two relatively mild motivating instructions: "These traits tend to be associated with happy marriages" or "These traits are predictive of happy middle-age." Typically we administered 20 traits with a positive motivation and 20 with a negative motivation. Because the two sets were synonyms, we simply subtracted each negatively-biased item from the corresponding positively biased item. The sum of these differences over the 20 pairs provided a simple index of motivational bias that could be administered along with other individual difference measures. Note that debriefing should be immediate and forceful.

Validation for this instrument was sought in several ways. Correlations with known indexes of distortion, for example, Self-Deceptive Enhancement (Paulhus, 1990), revealed substantial associations. Moreover, the predicted negative link with neuroticism was demonstrated. Predictable links to the other Big Five traits were also found.

CONCLUSION

Clearly all three experimental paradigms show promise for the experimental study of defense mechanisms. Each is consistent with some traditional aspect of defenses, although none of the three directly encompasses all the qualities attributed to psychoanalytic defenses, for example (Paulhus, Fridhandler, & Hayes, in press). What is not clear is the connection among these three and traditional approaches. I hope to address this topic at the next Gulf-Bridging Symposium on Defense Mechanisms.

DELROY
University
Vancouver

This
Symposi
in Niebo
testabilit
concept f
the pale o
symposiur
descriptio
construct
All th
example,
all permi
some em

Please send
British Col

6 Polish Psyc

REFERENCES

- Bargh, J. A. (1984). Automatic and conscious processing. In R. S. Wyer & T. K. Srull (Eds.), *Handbook of social cognition* (Vol. 3, pp. 1-43). Hillsdale, NJ: Erlbaum.
- Clark, M. S., & Isen, A. M. (1982). Towards understanding the relationship between feeling states and social behavior. In A. Hastorf & A. M. Isen (Eds.), *Cognitive social psychology* (pp. 73-108). New York: Elsevier/North-Holland.
- Ferguson, T. J., Rule, B. G., & Carlson, D. (1983). Memory for personally relevant information. *Journal of Personality and Social Psychology*, *44*, 251-261.
- Norman, W. T. (1963). Personality measurement, faking, and detection: An assessment method for use in personnel selection. *Journal of Applied Psychology*, *47*, 225-241.
- Paulhus, D. L. (1990). Measurement and control of response bias. In J. R. Robinson, Shaver, P., & Wrightsman, L. (Eds.), *Measures of personality and social psychological attitudes*. San Diego: Academic Press.
- Paulhus, D. L., & Baumeister, R. F. (1990). *Automatic and controlled self-presentation*. Unpublished manuscript.
- Paulhus, D. L., Fridhandler, B., Hayes, S. (in press). Psychological defense. In Briggs, S., Hogan, R., & Jones, W. (Eds.), *Handbook of personality psychology*, Academic Press.
- Paulhus, D. L., Graf, P., & van Selest, M. (1989). Attentional load increases the positivity of self-presentation. *Social Cognition*, *7*, 389-400.
- Paulhus, D. L., & Levitt, K. (1987). Desirable responding triggered by affect: Automatic egotism? *Journal of Personality and Social Psychology*, *52*, 245-259.
- Paulhus, D. L., & Lim, D. T. K. (1987). *Arousal and evaluative extremity: A dynamic complexity analysis*. Unpublished manuscript, University of British Columbia.
- Paulhus, D. L., & Murphy, G. (1987). *Disruption of impression management*. Unpublished study, University of British Columbia.
- Paulhus, D. L., & Suedfeld, P. (1988). Self-deception: A dynamic complexity model. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* New York: Prentice-Hall.
- Posner, M. I., & Snyder, C. R. R. (1975). Attention and cognitive control. In R. L. Solso (Ed.), *Information processing and cognition: The Loyola Symposium*. Hillsdale, NJ: Erlbaum.
- Qattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and the voter's illusion. *Journal of Personality and Social Psychology*, *46*, 237-248.
- Hamilton, P., Hockey, G. R. J., & Rejman, R. (1977). The place of the concept of activation in human information processing theory: An integrative approach. In S. Dornic (Ed.), *Attention and performance: VI* (pp. 463-468). New York: Academic Press.
- Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I Detection, search, and attention. *Psychological Review*, *84*, 1-66.
- Simon, H. A. (1986). Alternative visions of rationality. In H. R. Arkes, & K. R. Hammond (Eds.), *Judgment and decision making: An interdisciplinary reader* (pp. 97-113). New York: Cambridge University Press.