Why Intelligent Systems Should Get Depressed Occasionally and Appropriately

Charles Webster
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA 15261
cww@pogo.isp.pitt.edu

Abstract

Some researchers suggest that depression may be adaptive. For example, depression may provide an opportunity to assess our capabilities, learn from past failures, trigger personal change, and allocate activity away from futile goals. There are a variety of signature phenomena associated with depression, such as stable, global, and internal styles of failure explanation, a cognitive loop of failure-related rumination, lowered self-esteem and self-efficacy, and increased negative generalization and depressive realism. DEPlanner is presented, a simulated agent that adapts to failure in a simulated environment and exhibits eight targeted signature phenomena of depression.

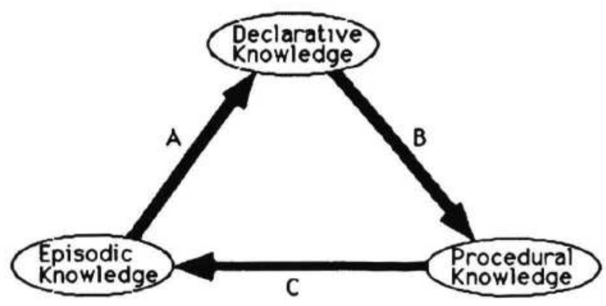
Introduction

Some types of depression in response to personal failure may be adaptive. Taylor (1989, p. 225) suggests that depression serves as a reality check, an opportunity to take "realistic stock of what one is and where one is going" and make "accurate assessment of his or her capabilities." Williams et al. (1988, p. 183) suggest that depression facilitates coping with long-term problems by "strategic access to previous problem-solving attempts." Flach (1974) and Gut (1989) argue that depression is a normal response to personal failure, and is necessary for personal change. Nesse (1991) hypothesizes that decreased mood reduces activity in situations where effort will not be rewarded, allocating energy away from bad investments. If depression can serve an adaptive function, then perhaps a simulated agent can be designed in such a way as to adapt to failure while exhibiting signature phenomena of depression. DEPlanner, based on ideas discussed in Webster et al. (1988), is such an adaptive agent.

This research was supported by the University of Pittsburgh Alzheimer Disease Research Center.

Overview of DEPlanner Simulation

DEPlanner is a nonlinear planner (Chapman, 1987) that generates sequences of instantiated STRIPS-style operators in response to initial and goal states. Since representing the social environment of a human is so complex, and since something like depression may be generally useful for adapting to some kinds of environmental change, planning and learning occur in a Blocksworld micro-world. DEPlanner's environment is a description of a set of blocks stacked on each other and scattered among locations on a table. Goals are conjunctions of targeted block configurations. A plan is a sequence of block movements. Successes have positive utilities, and failures negative utilities. Interactions between DEPlanner and its environment occur through the use of a problem generator that produces a stream of randomly constructed Blocksworld problems. The problem generator periodically changes internal rules that determine whether one block can be stacked on another block or location. A plan violating such a rule results in failure, and a deduction from DEPlanner's score. DEPlanner's task is to adapt in such a way as to maximize total utility accrued during interactions with the problem generator.



Information Flow in DEPlanner
Fig. 1

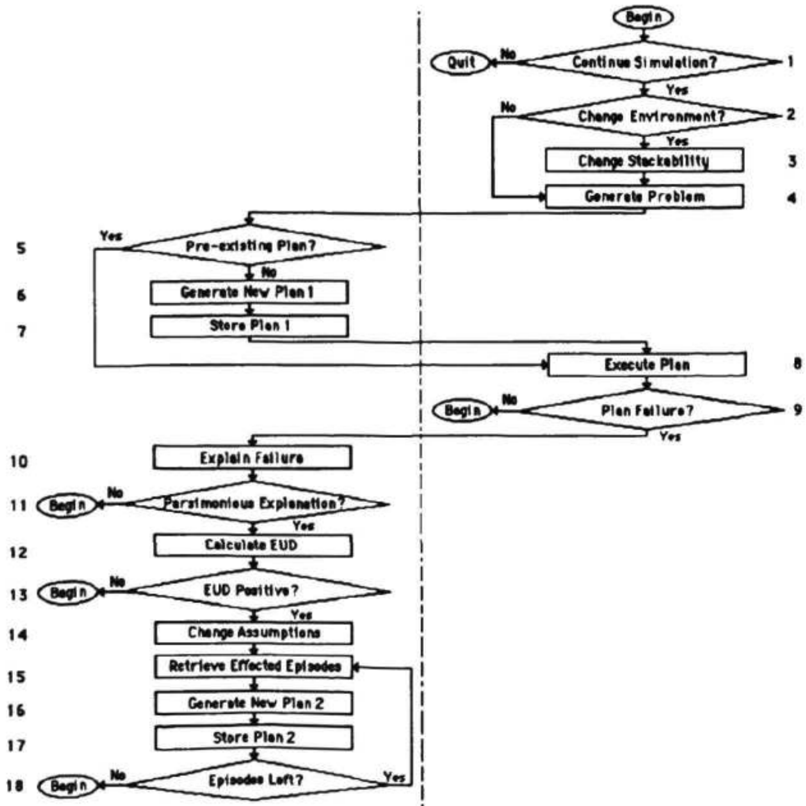
Like Soar (Rosenbloom et al. 1991) DEPlanner relies on declarative, procedural, and episodic representations. Fig. 1 depicts the flow of information among them. Declarative knowledge corresponds to representations of which block can be stacked on which block or location. (ASSUME (STACKONABLE B ?X)) means that block A can be stacked anywhere. Fig. 2 illustrates a plan operator. The "SENSE" preconditions match against the problem description provided by the problem generator, while the "ASSUME" precondition matchs against DEPlanner's revisable assumptions in declarative memory. Thus, changing declarative assumptions can result in the production of different plans.

```
#S(OP
  :NAME MOVE          ;move
  :VARS (?X ?Y ?Z)    ;x from y to z
  :PRECONDS
    ((ASSUME (STACKONABLE ?X ?Z))
        ;matches revisable assumption
     (SENSE (CLEAR ?X));matches problem description
     (SENSE (CLEAR ?Z));  "          "          "
     (SENSE (ON ?X ?Y))); "          "          "
  :ADD-LIST ((SENSE (ON ?X ?Z))
             (SENSE (CLEAR ?Y))
  :DEL-LIST ((SENSE (ON ?X ?Y))
             (SENSE (CLEAR ?Z))))
```

DEPlanner Operator
Fig. 2

Procedural knowledge consists of plans like ((MOVE B LOC1) (MOVE A B)) (along with appropriate preconditions and effects). Episodic knowledge is based on a time stamp, problem initial state and goals, the plan used, and outcome (success or failure).

DEPlanner's declarative knowledge changes in response to patterns of failure and success observed in episodic knowledge (A in Fig. 1). Procedural knowledge arises from the application of declarative knowledge (B in Fig. 1), creating a practice effect (DEPlanner becomes faster with experience). In order to change declarative knowledge on the basis of experience, the results of using procedural knowledge are recorded in the form of episodic knowledge (C in Fig. 1). When failures in episodic memory trigger changes in declarative memory, procedural knowledge must also change in order to remain a compiled version of declarative knowledge. DEPlanner's depression is the computationally expensive process of consulting episodic knowledge (retrieving past failures), modifying declarative knowledge (explaining patterns of failure), and compiling new procedural knowledge (creating new plans). Specialized control mechanisms suppress depression unless, and until, future benefits of depression are estimated to outweigh current costs.



Interaction Between DEPlanner (left) and Problem Generator (right)
Fig. 3

The flowchart in Fig. 3 depicts the flow of information and control between, and within, DEPlanner and its problem generator. The problem generator produces a problem (1-4, Fig. 3). DEPlanner returns a plan (5-7, Fig. 3). The problem generator determines whether the plan is a success (8-9, Fig. 3). And DEPlanner, if certain criteria are met, gets "depressed" (10-18, Fig. 3).

DEPlanner gradually increases its average response speed by reusing plans (5, Fig. 3). Plans can be stored (7, Fig. 3) after generation (6, Fig. 3) in response to problems posed by the problem generator, or they can be regenerated an stored (16 and 17, Fig. 3) under new assumptions in response to memories of old problems (15, Fig. 3). The problem generator periodically changes internal rules about which block can be stacked on which block or location (2 and 3, Fig. 3). If DEPlanner finds a set of assumptions that explains the observed pattern of recent failure (10 and 11, Fig. 3) it calculates a numerical heuristic Expected Utility of Depression (EUD) (12, Fig. 3). If EUD is positive (13, Fig. 3) then DEPlanner changes its assumptions (14, Fig.3) and retrieves previous problem-solving episodes now predicted to result in failure (15, Fig. 3). For each episode (18, Fig. 3) DEPlanner generates a new plan under new assumptions (16, Fig. 3) and stores it (17, Fig. 3).

Signature Phenomena of Depression as Design Constraints on DEPlanner

I used eight signature phenomena, culled from the research literature on depression, as a set of design constraints on the construction of DEPlanner: stable, global, and internal styles of failure explanation (Peterson & Seligman, 1984); a cognitive loop of failure-related rumination (Ingram, 1984); lowered self-esteem (Musson & Alloy, 1988) and self-efficacy (Bandura, 1977; Rehm, 1988); and increased negative generalization (Beck et al. 1979) and depressive realism (Alloy & Abramson, 1988). My task was constructing a simulated agent that adapts to changes in a simulated environment, and exhibits as many signature phenomena of depression as possible.

The first three phenomena (stability, globality, and internality) concern the kinds of events that are most likely to trigger a depression. It is not failure itself, but rather explanation for failure that serves as a trigger. Failure attributions that are stable (continuing to be true in the future), global (affecting many important goals), and internal (having been avoidable) are most likely to precipitate depression (Peterson & Seligman, 1984). Each dimension can be mapped to a different aspect of the formula (Eq. 1) used to trigger or suppress DEPlanner's analog of adaptive depression.

If DEPlanner finds a parsimonious explanation for a recent pattern of failures, in terms of a change in assumptions, DEPlanner uses Eq. 1 to calculate the Expected Utility of Depression (EUD). If EUD is positive for forgoing current opportunities (PGI times PAU), but preparing for future opportunities (PSI times PMU), then DEPlanner gets "depressed." If the EUD is negative then DEPlanner "shrugs off" failure and moves on.

PMU = average Plan Marginal Utility per unit time
(difference in utility accrued per unit time between having and not having a plan)

PSI = average Plan Soundness Interval
(length of time over which a plan will prove useful)

PAU = average Plan Achievable Utility per unit time
(utility accrued per unit time from current repertoire of plans)

PGI = average Plan Generation Interval
(length of time required to generate a plan)

$$EUD = (PMU * PSI) - (PAU * PGI) \qquad \text{Eq. 1}$$

The parameters in Eq. 1 can be associated with the three dimensions of failure attribution in such a way that increasing stability, globality, or internality correspond to increasing EUD and likelihood of depression. The more stable an environment, the longer a period of time over which the benefits of a precomputed plan can be amortized. Therefore stability is associated with the PSI parameter. The more globally important an environmental change, the larger the number of important affected goals. Both PMU and PAU can be associated with globality (PMU positively, and PAU negatively) because they both vary with the total utility affected (in PMU's case) or unaffected (in PAU's case). As for internality, an internal failure attribution is one in which we believe that we could have avoided failure if we had pursued an alternative course of action. In DEPlanner's case an alternative course of action is an alternative plan. The less effort or time required to generate an alternative plan, the more internal should be the attribution. Conversely, impossible plans, which have infinitely long generation times, should result in large PGI's. Therefore internality is mapped to 1.0 / PGI.

When DEPlanner's depression ensues, five more signature phenomena appear: the cognitive loop of failure-related rumination, lowered self-esteem and self-efficacy, and increased negative generalization and depressive realism.

Depressed people experience a cognitive loop of failure-related thoughts that distract them from normally enjoyable activities (Ingram, 1984). During the cognitive loop people think about past failures, construct possible explanations, and consider future implications. When the cognitive loop recedes, people often make important changes in their lives (such as changing their expectations or focusing on different goals). In DEPlanner the cognitive loop consists of retrieving failures from a memory of past problem-solving episodes (accounting for William et al.'s (1988, p. 183) "strategic access to previous problem-solving attempts."), searching for failure explanations (accounting for Taylor's (1989, p. 225) "accurate assessment of his or her capabilities"), and generating new plans under new assumptions. Thus, DEPlanner's analog of the cognitive loop is at the core of DEPlanner's adaptive ability.

William James defined self-esteem to be total successes divided by total pretensions (James, 1890). In DEPlanner's case, self-esteem is the total utility of achievable goals divided by the total utility of all goals in episodic memory. Self-esteem ranges from 0.0 (no goals are possible) to 1.0 (all goals are possible).

Self-efficacy, roughly the subjective probability that a goal can be accomplished, is thought to drop during depression (Rehm, 1988). In DEPlanner self-efficacy is defined in a similar fashion to self-esteem, except that utility is not taken into account: total number of attainable goals divided by total number of goals.

People who are vulnerable to depression are more likely to arrive at negative self-deprecating generalizations in response to small setbacks (Beck et al. 1979). For example a mother whose child complains of a cold breakfast may decide she is a bad mother, rather than she is not a perfect cook. DEPlanner's assumptions, about which block can be stacked on which block or location, are indexed in a hierarchy, with more general assumptions toward the top, and more specific assumptions toward the bottom. (Retracting (ASSUME (STACKONABLE B ?X)) is more general than retracting (ASSUME (STACKONABLE B A)) because retracting (ASSUME (STACKONABLE B ?X)) effectively retracts all assumptions beneath it in the hierarchy.) Given a set of successful and failing past problem-solving attempts, DEPlanner finds positions in the generalization hierarchy that predict the pattern of successes and failures. The more general a failure attribution, the more "ASSUME" preconditions are unsatisfied, and the longer DEPlanner spends precomputing new plans. Thus DEPlanner's behavior is consistent with the correlation between human negative generalization and depression. DEPlanner's measure of negative generalization is the number of retracted assumptions divided by the total number of possible assumptions.

The depressive realism phenomenon is particularly problematic for other theories of depression. In some ways, mildly depressed people appear to be more accurate information processors than nondepressed people (Alloy & Abramson, 1988). For example, nondepressed people seem to over-estimate their chances of success at a variety of tasks, while depressed people are relatively more accurate. A simple measure of DEPlanner's depressive realism is total actual goal utility (according to correct assumptions available in the problem generator) divided by total predicted goal utility (according to DEPlanner's own assumptions and plans). During depression, DEPlanner's measure of depressive realism increases because DEPlanner's assumptions become more accurate.

Comparing DEPlanner's Behavior to Signature Phenomena of Depression

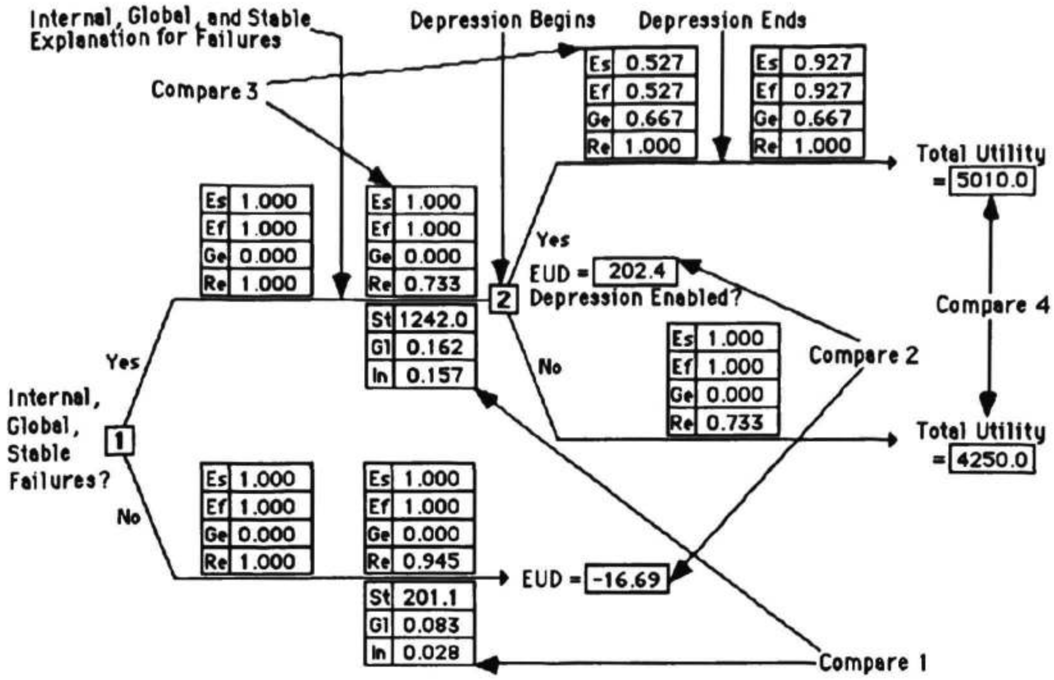
Cognitive modellers aspire to detailed statistical comparisons between the behavior of their simulations and human subjects. However many simulations, especially those that are the first to model a psychological process, do well to qualitatively match a set of signature behaviors obtained from the research literature. DEPlanner is in this latter category. In order to confirm that DEPlanner behaves according to its intended design, an experiment (Fig. 4) was devised.

The value of each number, or change in value, is not important for the purpose of this qualitative assessment. It is the order and direction of change that is relevant. Each of the two decision boxes corresponds to an experimental manipulation. Box 1 corresponds to the presence or absence of a stable, global, and internal environmental change. Box 2 corresponds to the presence or absence of DEPlanner's analog of adaptive depression.

DEPlanner's problem generator has "knobs" that influence the frequency of change (stability), the frequency and total utility of goals affected (globality), and length of time required to construct plans (internality). In order to set these parameters, DEPlanner's assumptions, plans, and past problem-solving attempts are accessed, but not changed. The resulting statistics are used to determine which rules to change in the problem generator, and how often.

Beneath the paths leaving box 1 are tables of numbers representing stability (St), globality (Gl), and internality (In). They are higher on the upper path (see "Compare 1", Fig. 4), consistent with higher stability, globality, and internality. Here the Expected Utility of Depression (EUD) is positive and triggers DEPlanner's adaptive depression (see "Compare 2", Fig. 4). In the low stability, globality, and internality condition EUD is negative (but almost positive) and DEPlanner suppresses its depression.

Box 2 corresponds to enabling (upper path) or disabling (lower path) DEPlanner's adaptive depression. Where adaptive depression is enabled, we see the signature phenomena of decreased self-esteem (Es) and self-efficacy (Ef), and increased negative generalization (Ge) and depressive realism (Re) (see "Compare 3", Fig. 4). Self-esteem and self-efficacy



Effects of Environmental and Architectural Change on Signature Phenomena
Fig. 4

have identical values because the current version of DEPlanner's problem generator assigns the same utilities to all problem-solving attempts (5.0 for successes and -5.0 for failures). These numbers would diverge if different problems had different utilities.

Self-esteem and self-efficacy drop because the total utility of unthreatened goals drops. Negative generalization increases in this particular case because DEPlanner retracts (ASSUME (STACKONABLE B ?X)), effectively retracting a large number of assumptions. Correcting DEPlanner's assumptions increases DEPlanner's measure of depression realism.

Eventually all plans that can be precomputed, have been precomputed, and self-esteem and self-efficacy move almost, but not quite, back to their original values. They fail to reach 1.0 because some plans are impossible in light of the change in assumptions. For instance, a goal set that includes (ON B A) is impossible because (ASSUME (STACKONABLE B A) has been retracted.

In either condition, depression-enabled or depression-disabled, the DEPlanner simulation continues until the next change in the environment, whereupon the value of total utility accrued is calculated (see "Compare 4", Fig. 4). When DEPlanner is allowed to get "depressed" the total utility is 5010.0. When DEPlanner's depression is disabled the total utility is 4250.0. Thus DEPlanner achieves a 18 per cent higher total utility in the depression-enabled condition. DEPlanner exhibits analogs for signature phenomena of depression <u>and</u> shows how they may be generated by a functionally useful mechanism.

Conclusion

To become a better model of depression, DEPlanner requires redesigning. Relevant issues include:

1. A more plausible model of goal-driven problem solving and learning should replace the nonlinear planning algorithm. Computer models of knowledge acquisition and compilation in human students will be relevant (e.g., VanLehn et al. 1992).
2. The domain should be changed to reflect concerns of depressed people. Representing changing career and family circumstances will require a more expressive knowledge representation.
3. DEPlanner should be better motivated in terms of concepts like self-esteem and self-schema, as well as general theories of emotional information processing.
4. While DEPlanner is not a model of abnormally triggered, sustained, or intense depression, DEPlanner might be "broken" or "lesioned" in order to provide such an account.
5. The effects of external interventions, analogous to cognitive therapy for depression (Beck et al. 1979), should be modeled with respect to reducing likelihood, intensity, or duration of DEPlanner's depression.
6. Other kinds of cognitive dysfunction, such as Alzheimer Disease, can masquerade as depression, and visa versa (Caine, 81; Merriam et al. 1988). Perhaps DEPlanner can be broken in different ways in order to account for different but related syndromes.

7. DEPlanner uses a heuristic to modulate its cognitive loop. A sophisticated control regime should be based on an explicitly normative metareasoning approach (Horvitz et al. 1991).

8. DEPlanner should be situated with respect to existing systems in the space of cognitive architectures, such as case-based reasoners and explanation-based learners.

9. Extensive sensitivity analysis will be required to understand the complex interactions among DEPlanner's parameters, and between DEPlanner and its environment.

Nevertheless, the current version of DEPlanner is consistent with a relatively large set of signature phenomena associated with depression, and the hypothesis that some forms of depression may be adaptive. Rudimentary computational mechanisms can be assembled into a model of depression that explains a large set of previously unrelated signature phenomena.

Acknowledgements

I thank Kurt VanLehn, Greg Cooper, Gordon Banks, Johanna Moore, Richmond Thomason, Martha Pollack, and two anonymous reviewers, for helpful comments and useful criticisms. Thanks also to Steve Small for access to computing facilities at the Cognitive Modelling Laboratory.

References

Alloy, L. & Abramson, L. (1988). Depressive Realism: Four Theoretical Perspectives. In L. Alloy (Ed.), *Cognitive Processes in Depression.* New York: Guilford Press.

Bandura, A. (1977). Self-Efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, 84, 191-215.

Beck, A., Rush, A., Shaw, B. & Emery, G. (1979). *Cognitive Therapy for Depression: A Treatment Manual.* New York: Guilford Press.

Caine, E. (1981). Pseudodementia: Current Concepts and Future Directions. *Archives of General Psychiatry*, 38, 1359-1364.

Chapman, D. (1987). Planning for Conjunctive Goals, *Artificial Intelligence*, 32, 333-377.

Flach, F. (1974). *The Secret Strength of Depression.* New York: Lippincott.

Gut, E. (1989). *Productive and Unproductive Depression: Success or Failure of a Vital Process.* New York: Basic Books.

Horvitz. E., Cooper, G. & Heckerman, D. (1989). Reflection and Action Under Scarce Resources: Theoretical Principles and Empirical Study. *11th Int. Journal Conference on Artificial Intelligence.*

Ingram, R. (1984). Toward an Information-Processing Analysis of Depression. *Cognitive Therapy and Research*, 8, 443-478.

James, W. (1890). *The Principles of Psychology, Vol. 1.* Henry Holt & Co.

Merriam, A., Aronson, K., Gaston, P., Wey, S., & Katz, I. (1988). The Psychiatric Symptoms of Alzheimer's Disease. *Journal of the American Geriatric Society*, 36, 7-12.

Musson, R. & Alloy, L. (1988). Depression and Self-Directed Attention. In L. Alloy (Ed.), *Cognitive Processes in Depression.* New York: Guilford Press.

Nesse, R. (1991). What is Mood for? *Psycoloquy*, Vol. 2, Issue 9.2.

Peterson, C. & Seligman, M. (1984). Causal Explanations as a Risk Factor for Depression: Theory and Evidence. *Psychological Review*, 91, 347-374.

Rehm, L. (1988). Self-Management and Cognitive Processes in Depression. In L. Alloy (Ed.), *Cognitive Processes in Depression.* New York: Guilford Press.

Rosenbloom, P., Newell, A. & Laird, J. (1991). Toward the Knowledge Level In Soar: The Role of the Architecture in the Use of Knowledge. In K. VanLehn (Ed.) *Architectures for Intelligence*, Hillsdale, New Jersey: Lawrence Erlbaum.

Taylor, S. (1989). *Positive Illusions: Creative Self-Deception and the Healthy Mind.* Basic Books.

VanLehn, K., Jones, R., & Chi, M. (1992). A Model of the Self-Explanation Effect. In M. Posner (Ed.), *Journal of the Learning Sciences*, 2, 1, 1-59.

Webster, C., Glass, R. & Banks, G. (1988). A Computational Model of Reactive Depression. *Proc. 10th Annual Conf. Cognitive Science Society*, Montreal, Canada: Lawrence Erlbaum.

Williams, J., Watts, F., MacLeod, C. & Mathews, A. (1988). *Cognitive Psychology and Emotional Disorders.* New York: John Wiley & Sons.