George Ainslie
Veterans Administration Medical Center, Coatesville

Jefferson Medical College

ABSTRACT: Freud was the first author to conceive internal motivational conflict in economic terms. Although behaviorists have often rejected his concepts because the findings that gave rise to them were based on subjective methods, they are largely compatible with behavioral data on motivation, and indeed predicted by Herrnstein's matching law. Psychoanalysis is much closer to behavioral than to cognitive psychology, which does not conceive self-contraditory behavior as a motivational problem.

Picoeconomics ("micromicroeconomics") is the use of behavioral economic principles to analyze motivational conflict within the individual organism (Ainslie, 1986 and forthcoming). It is especially applicable to situations where the organism seems divided, that is, where its behavior appears to be self-defeating or irrational. Many of its concepts can be traced to the first author to deal with such situations comprehensively, Sigmund Freud.

Freud had such a defining effect on our understanding of irrational behavior that it is hard to imagine what this branch of knowledge would have been like without his influence. Since his death his theories have been widely attacked by behaviorists, and widely ignored by the other major school that deals with irrationality, cognitive science. However, this wholesale discard of his contributions is unwarranted. I will reexamine some of his major concepts about motivational conflict in the light of recent behavioral research.

The issue of whether psychoanalytic therapy is effective is only marginally relevant to the question of whether these concepts are valid. Amid the number of studies that detect no difference between analytic therapy and other psychotherapies, or no therapy, there are occasional reports that find analytic therapy superior, even in such an unlikely population as drug addicts (Luborsky et al., 1985). Whether such reports should be believed in the face of a number of negative findings has become a highly technical question (Brown, 1987). Our discussion does not turn on its answer. If psychoanalysis turns out to be helpless its concepts could nevertheless be valid, and if it turns out to be effective this might be for extraneous reasons.

Freud himself was a pessimist about therapy, and regarded himself principally as a scientific observer. Thus the behaviorists' challenge of his views on motivation goes to the heart of his contribution. However, a distinction should be made between behaviorism as a philosophy of science, often called "radical behaviorism," and behaviorism as a body of empirical findings. Behaviorism as a philosophy of science is inimical to Freud's methods. He did not base his conclusions on controlled experimentation, and usually used subjective data: patient self-reports, letters or published reports about patients he had never seen, myths, and even veiled introspections. However, radical behaviorists have not demonstrated that subjective data are less valid than other kinds of observations, only that they offer the observer a greater temptation to self- deception (Ainslie, 1985). Properly used, such data

#### **AINSLIE**

adequately serve other branches of science like linguistics. Behaviorism as a body of research is not incompatible with Freud's findings; quite the contrary, it is extremely useful in interpreting these findings, as we shall see.

The greatest barrier between Freud and modern scientists does not arise from his observations but from the theoretical structure he built on them. He had not one but several explanations of why people develop such complicated mental lives, and all of them were couched in the terms of Victorian neurology, hydraulics, electrostatics, and other nascent fields, the languages of which sound bizarre to the modern ear. Even when he used bald, homespun terms like the "1" and the "it" and "dwelling on," his translator, Ernest Jones, rendered his ideas for English readers in Latin and Greek terms like ego, id, and cathexis. As a result, Freud's theoretical explanations sound like nothing so much as a cabalistic religion. There have been many psychoanalysts who have found them a barrier to practical discussion (e.g., Klein, 1973; Michels, 1983).

Some of the problems presented by this exotic theoretical structure, usually referred to as the "metapsychology," are trivial. Freud meant many of his models as analogies or illustrations. Even where he literally believed in a certain underlying neurological configuration, he was summarizing actual observations and could be seen by a sympathetic reader as defining the supposed neurological mechanism functionally. But when we strip away the imagery to see what he "really meant," we find that he really meant several different hypotheses that are neither articulated with nor cleanly opposed to one another.

Freud's main concern was impulses: motivated but seemingly irrational behaviors that undermine a person's best interests as recognized by his conscious intentions. Having publicized if not discovered the extent of human impulsiveness, he was at some pains to find a scientific explanation. His first hypothesis (with Josef Breuer, 1895) invoked hypnosis, itself a phenomenon lacking a fundamental explanation but one that was at least well known: He suggested that during periods of great emotional stress, a person entered a "hypnoid state," and was thereby programmed to repeat whatever it was he experienced in that state whenever the happenstance of stimuli reminded him of it. He rarely referred to this mechanism in his later, more comprehensive theories, but the idea of a compulsion to repeat has remained in the analytic literature down to the present.

At about the same time, Freud was developing a model in neuronal terms, although it behaved not like any known neural system, but rather like an electrical capacitor (1895): Stimuli or perceptions that were too intense to be tolerated were diverted into the unconscious, which thereafter tended to discharge them either back into consciousness or into motor activity. The unconscious would then function as an autonomous center of behavioral decision making, made shortsighted by its separation from consciousness. When it discharged its energy into motor activity, it created impulses.

Somewhat later (1911) he proposed that there were two principles of mental functioning, a pleasure principle and a reality principle. The impulsiveness of the former stemmed from an overvaluation of immediate results. Over time, a more farsighted principle arose from the pleasure principle through learning: "The substitution of the reality for the pleasure principle implies no deposing of the pleasure principle, but only a safe-guarding of it. A momentary pleasure, uncertain in its results, is given up, but only in order to gain along the new path an assured pleasure at a later time (p. 223)." This implies that the conflict between impulse and control is not between separately motivated decision-makers but only between ways of discounting delayed events.

Finally (1920), he returned to the idea of two autonomous centers of motivation (id and ego), although these were not meant to be anatomic locations. Since they were autonomous

there might be choices that gave the id pleasure but gave the ego unpleasure. As in his second model, the autonomy of the impulsive center came from this unconsciousness; in principle its influence could be eliminated by making the person conscious of the relevant motives.

Meanwhile, in other writings, Freud was developing a third factor that was more antiimpulsive than the ego, sometimes to the point of maladaptive rigidity: the superego. He
conceived this conscience-like factor to arise through a process like classical conditioning,
but its force was often disproportionate to the punishments that might have led to
conditioning. He ascribed this distortion to an expectation of punishment that the person
formed a priori because of his wishes, especially Oedipal wishes. The superego's
independence from other motives was again explained by its being at least partially
unconscious.

Freud was clearly starting from clinical observations and casting about for a theoretical model that would fit them. He was not the first to describe a relatively rational self that is pressured on one side by impulses and on the other by some kind of conscientiousness—Plato had described much the same model (1961; others in Ainslie, 1982). Freud's contribution was his broad systematic attempt to account for all seemingly irrational and involuntary behaviors with this model. Furthermore, with the exception of his early idea about hypnoid states, he did so without abandoning the strict assumption that all behavior was motivated, and hence reducible in principle to some kind of economics. However, his economic model using the hypothetical currency of libido did not integrate his other models. What he left was a somewhat related body of astute observations on motivational conflict, with sketches of an attempt to derive these from the Victorian understanding of biology. The question psychoanalysts and behaviorists have debated ever since is whether the metapsychology can be repaired and built upon, or whether it must be abandoned in favor of a radically different approach.

The answer I propose is that a radically different approach—behavioral psychology—is not nearly so incompatible with Freud's as the partisans of both sides maintained during the 1950's and 60's, and can be used to repair and build upon Freud's formulation. From the latest discoveries of basic behavioral research, it is possible to derive a motivational rationale for most although not all of the major concepts that evolved from Freud's observations, and to reconcile them with the ideas of motivation current in other sciences. Moreover, it will be possible to suggest some solutions to long-standing puzzles in psychology, such as the relationship of aversive to rewarding events, the motivational basis for the competition between fantasy and reality, and the constraints on game-like or "process" rewards.

The link with modern motivational science cannot be made without some damage to Freud's system. His most significant error seems to have been that in constructing his final model of the ego, id, and superego, he relied upon his second hypothesis about the pathogenesis of ambivalence rather than his third: He developed the theory that unpleasant stimuli are diverted to a different functional location and are thus made partially autonomous from the person's general motives, rather than following his simpler idea that, in evaluating future prospects, a person by nature discounts delayed events sharply.

Although repression and a body of unconscious material demonstrably exist in everyone, this is not evidence that repression is the fundamental cause of impulsiveness. On the contrary, repression often seems to be an impulsive act in its own right, in that it accepts a greater pain in the future in return for avoiding a smaller pain in the present. At other times it represents an attempt to control impulses, as we shall see shortly. Thus repression is frequently seen in the vicinity of impulses, and is a natural object of our suspicion; but although it is a weapon in the war of motive against motive, no finding from any science of

### **AINSLIE**

motivation suggests that it could be a cause of the war. The theory of repression itself entails no exception to the utilitarian model in which the greater motive always prevails over lesser ones.

On the other hand, there is good evidence that both lower animals and people spontaneously perceive the value of delayed events as markedly lower than that of imminent ones. Not only that, but the shape of the discount curve as a function of delay is highly concave upwards, best described not with the exponential curves by which financial discounting occurs, but by hyperbolae. That is, according to Herrnstein's matching law, the raw, uncorrected value of an event is inversely proportional to its delay (Herrnstein, 1981). This is the same proportionality of perception which holds true of many other dimensions like brightness and loudness (Gibbon, 1977), but when it occurs in the judgment of temporal relationships some startling consequences ensue.

Hyperbolic discount curves drawn from two alternative goods, one smaller but available earlier than the other, will predict a period of temporary preference for the smaller good as a function simply of elapsing time (Figure 1a). The heights of the curves are proportional to the heights of their goods only when both goods are relatively distant; but at that point they still represent an incentive to choose "objectively" between the goods. However, unless this choice includes some means of forestalling the subsequent change of preference, this decision will come to naught (Ainslie, 1975).

These highly bowed discount curves have been observed in every organism so far studied (deVilliers and Herrnstein, 1976; Vaughan and Herrnstein, 1985). However, people are also widely observed to choose "objectively," a discrepancy which has led to a heated debate about whether hyperboloid curves are the fundamental discount function or not (Logue, 1988). Freud's concept of the conflict between pleasure and reality principles suggests a way to integrate these observations.

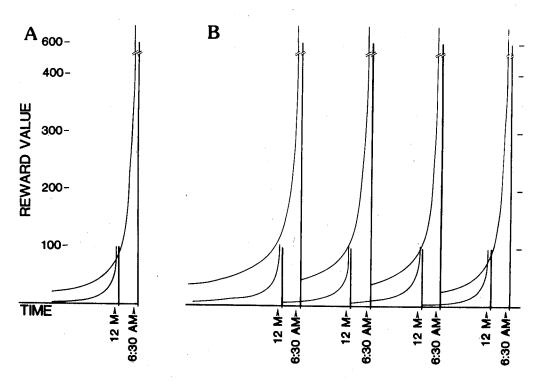
A hyperboloid discount function predicts that a person cannot simply choose between some kinds of alternatives, because he cannot weigh them against each other to produce a stable preference. The incompatible motives stay in competition. Only the short term motive ever wins a solid victory, for by the time it is preferred the long term motive has lost its chance; but hundreds of successors to the current short term motive must continue to do asymmetrical battle with the farsighted motives. This competition is identical to that of Freud's two principles of mental functioning. In the inexperienced individual, the curves motivate shortsighted mechanisms to get immediate goods, as if according to the pleasure principle. As he gets more experience he will learn farsighted devices which will be more weakly motivated than the shortsighted mechanisms but which will have the advantage of being dominant first. If they can forestall the impending dominance of the pleasure principle they will increase the individual's objective income over time, thereby constituting a reality principle to "safeguard" the goals of the pleasure principle just as Freud said.

Seen in this framework, the two principles do not consist of separate organs or centers of motivation, but only of rewards that cannot simply be weighed against their competitors, and of the operants that these rewards shape to conduct the competition among them. These entities might be called interests, since they are constituted within the person much like interests within a legislature: Non-dominant interests maintain their integrity and some degree of power insofar as they are likely to become dominant in the future.

The devices learned by interests to undermine each other look a lot like defense mechanisms. Four kinds are logically possible: Each has been described both in the psychoanalytic literature and by subjects who are interviewed about their self-control strategies (Ainslie, 1982 and forthcoming):

- (1) An interest can set in motion forces outside of the mind that will have a committing effect on other interests. Antabuse, which renders alcohol sickening for a period after it is taken, is a prime example of such a device. An adolescent who fears a major impulse commonly commits a more minor offense first so as to "ask for controls." People often join groups, ranging from Weight Watchers to the paratroops, which will exert social pressure in a particular direction. This device is usually used by long range interests against short range ones, but people in the grip of a short range interest may sometimes do things like discard the number of their AA contact or alienate a parent to deny resources to the long range interest, lest it reassert itself during the current impulsive episode.
- (2) A long range interest may forestall a short range one by arranging for the person not to receive information about the availability of imminent rewards. Conversely, a short range interest may hide from a long range one by not collecting or processing information about the nature of an impulsive act; the point of such concealment will be clearer when we have discussed the fourth device. Psychoanalysis calls information-controlling tactics by either interest suppression, repression or denial.
- (3) Since the major emotions like anger, love and fear seem to have some intrinsic momentum, short range interests that involve such emotions will be somewhat protected, once under way, from intervention by long range interests. However, a long range interest may forestall such a short range interest by preventing the early development of the relevant emotion, a device psychoanalysis calls isolation of affect, or by cultivating a contrary emotion, a process called reversal of affect of reaction formation.
- (4) The most powerful, yet most flexible move open to the long range interest is to define classes of similar behaviors, such that the person's choice in the present case provides him the best information predicting how he is apt to choose in similar cases in the future. When the individual's current choice sets a precedent, the reward at stake is not only the literal outcome of his current choice, but also his expectation of getting the whole class of larger-later versus smaller-earlier rewards.

In terms of the hyperbolic discount curves, defining individual choices as precedents transforms a choice of two individual events into a choice between the summed curves of two whole series of events; and since all but the immediate events are seen at some distance, the net effect of this change is to add motivation to the long range interest (Fig. 1b). This perceptual change is the basis of the personal rule, the backbone of what the Victorians called force of will and psychoanalysts refer to as the compulsive defenses. The countermove by a short term interest is either to distinguish the circumstances of the current choice from the criteria of the rule, the lawyerly process of finding loopholes called rationalization, or to obscure the information needed to judge whether a precedent was involved or violated, which is the case of repression in the service of the short range interests that I have already discussed.



#### FIGURE CAPTION

Fig. 1. Hypothetical values of staying up late versus going to bed, as a matching-law function of time (ValueAmount / Delay). At midnight (arrows) the person must choose between staying up (fun worth 100 points at a mean delay of an hour) and going to bed (comfort at 6:30 AM worth 600 points at a mean delay of an hour): A. for this night only, and B. for four consecutive nights. The curve for the delayed satisfaction becomes relatively higher than that for staying up as curves for subsequent days are added in.

The use of personal rules transforms a motivational conflict from being literally one of successive preferences to a simultaneous struggle between principles and exceptions: A short range reward may be constantly available and yet be preferred only when some distinction detaches it from a governing principle. Because of this transformation, the matching law can account not only for sequential conflicts of the Ulysses-and- Sirens type, but also for the struggles between simultaneous motives about which people more commonly complain.

Thus Freud's idea of impulsiveness as a product of discounting future events is supported by modern behaviorism, and a necessary property is added: the sharp upward concavity of the discount curves that generate temporary preferences for poorer alternatives. The consequences of this tendency to form temporary preferences are very much the same

defense mechanisms that he and his daughter Anna described. The suggestion of later writers that some defense mechanisms are healthy and even necessary (Haan, 1963; Vaillant, 1976) is also vindicated.

The predictions of this model go a great deal further, confirming some of Freud's hypotheses and correcting others. For instance, the evaluation of current choices as precedents represents a mechanism for the third force in Freud's structural model, the superego; however, it does not rely on the Oedipal conflict, which Freud believed necessary to account for why small incentives could sometimes occasion intense motives. The genuine importance of small precedents in this method of impulse control explains the potency, first recognized by Freud, of trivial situations that would otherwise seem to be mere symbols.

The need for stimulus patterns outside of the person's control to serve as criteria for personal rules explains some of the transformations that psychoanalysts have also called defense mechanisms (Ainslie, 1982): identification with or introjection of other people, displacement of meaning from one object onto another, and sublimation from one set of goals to a more subtle one. The investing of objects with importance by making them the criteria for personal rules is cathexis, the constraints on which are well defined for the first time in the temporary preference model. However, Freud's idea that cathexis and the defense mechanisms are built of psychic energy (motivation) rather than being themselves motivated behaviors is not supported, nor is his belief that the motivation needed for them has to exceed the motivations for the processes controlled. The motivation needed actually depends on the leverage that can be obtained from early dominance, or from summation with other outcomes for which it is a precedent.

The temporary preference concept of impulses also permits systematic hypotheses about areas that Freud only began to explore. For instance, Freud was impressed by people's power of fantasy, and felt a need to explain why people generally become less involved in fantasy as they grow up. He attributed this change to "the instincts of self-preservation" (1915); but in fact the fear of physical harm does not seem to be what limits people's fantasy life in modern society. Rather the matching law's hyperbolic discount curves suggest that in the many areas where we have arbitrary control over our rewards, we will tend to pace our consumption of them inefficiently unless we bind ourselves to criteria for self-reward that are outside of our arbitrary control. That is, if we rehearse a romance or adventure entirely in our heads, our overvaluation of early versus late reward will induce us to come to the high point earlier and earlier, so that the payoff over time does not repay whatever effort was needed to set up the fantasy (Ainslie, 1986). By this rationale, the motivational value of reality is as a discipline which restrains our own impulsive preference for the wastefully rapid consumption of reward.

Freud often suggested that emotions like fear and pain are not entirely aversive but are the objects of ambivalence. He ascribed such ambivalence to competing life and death wishes, but a simpler explanation is that such emotions are very briefly rewarding, rapidly satiating, rapidly recovering rewards (Ainslie, 1987a). Such rewards can command immediate attention- directing behavior because of the hyperbolic discount curve, but because they are brief they have much less effect on longer latency processes like muscle movement. Thus they can be both vivid and aversive.

These illustrations show that it is practical to outfit Freud with behaviorist underpinnings. The new relationship of behaviorism with economics can be enriched by rereading the first theorist to propose an economic model of psychic life. Psychoanalysis in turn can gain a bridge to the other sciences, a bridge that was energetically maintained by Freud but largely abandoned by his successors after the failures of direct experimental

### **AINSLIE**

demonstrations of his theories.

Is this collaboration apt to be of any interest to cognitive science, thus far a strictly noneconomic discipline? The boundaries of cognitive science are indistinct, but writings with "cognitive" in their title rarely say anything about motivation. Even when dealing clinically with symptoms of self- destructive behavior, the cognitive article focuses on problems of perception or logic, as if its subjects were trapped in a puzzle from which they would unambivalently escape, were they not deficient in some kind of information-processing skill.

The neglect of motivation will probably limit the ability of the cognitive method to deal with ambivalence. Much of the puzzle in which a person can become trapped seems to be imposed not by the environment but by the potential of his own motivation to fluctuate over time. Of course, this situation itself is often the subject of cognitions (Ainslie, 1987b). The cognitions that we understand the least, and those that especially limit the person who is trapped in self-defeating behavior are those that integrate, or fail to integrate, successive motivational states. To achieve the kind of consistency that society calls rational, a person must become adept at budgeting his own future rewards so that imminent but inferior goods rarely dominate his motivation. This kind of cognition is an order of magnitude more difficult than the logic which is now the staple of cognitive therapists' teaching.

Messer and Winokur have said that psychotherapists' views of human nature can be divided into the comic and the tragic (1980). The comic view is that man is perfectible, or at least clever enough to leave his shortcoming behind. The tragic view is that he is innately perverse. Behavior therapists have generally held the comic view, psychoanalysts the tragic. The same division occurred earlier in philosophy—the Hobbesian view of total depravity against Rousseau's noble savage; and in religion—for instance in debates about whether mankind was corrupted by the original sin of Adam.

In cognitive science the comic view is in the ascendant. There is more than a little of Rousseau in the view that self-defeating behavior is at its root just an ignorant mistake. But the matching law supports Freud in his Hobbesian view: Short- sightedness and impulsiveness are intrinsic to the way we see the world. There turns out to be a scientific basis for the myth of original sin: What Adam and Eve did in the Garden of Eden was to come upon the exponential discount curve of rational judgment and swing upon it, bending it into the hyperbolic curve that has generated self-defeating behaviors ever since.

This, too, can be grist for the mill of cognitive science. But the product is not apt to be a solution in the sense that puzzles have solutions; we will have to settle for clarification of yet another immutable limitation of human nature. Freud was the bearer of a tragic message, but it is one that withstands the light of modern science and becomes clearer in it.

#### REFERENCES

Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82, 463-496.

Anislie, G. (1982). A behavioral economic approach to the defense mechanisms: Freud's energy theory revisited. *Social Science Information*, 21, 735-779.

Ainslie, G. (1985). Behavior is what can be reinforced. The Behavioral and Brain Sciences, 8,53-54.

Ainslie, G. (1986). Beyond microeconomics: Conflict among interests in a multiple self as a determinant of value. In J. Elster (Ed.), *The Multiple Self*. Cambridge, England: Cambridge University Press.

- Ainslie, G. (1987a). Aversion with only one factor. In M.L. Commons, J.E. Mazur, J.A. Nevin, & H. Rachlin (Eds.), Quantitative Analyses of Behavior V: The Effect of Delay and of Intervening Events on Reinforcement Value. Hillsdale, NJ: Erlbaum.
- Ainslie, G. (1987b). Self-reported tactics of impulse control. The International Journal of the Addictions, 22, 167-179.
- Ainslie, G. (forthcoming). Picoeconomics: The Interaction of Successive Motivational States within the Individual. Cambridge, England: Cambridge University Press.
- Brown, J. (1987). A review of meta-analyses conducted on psychotherapy outcome research. Clinical Psychology Review, 7, 1-23.
- deVilliers, P. & Herrnstein, R. (1976). Toward a law of response strength. Psychological Bulletin, 83, 1131-1153.
- Freud, S. (1895). Project for a scientific psychology. In J. Strachey & A. Freud (Eds.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (pp. 295-341). London: Hogarth, 1956. Vol. 1.
- Freud, S. (1911). Formulations on the two principles of mental functioning. In J. Strachey & A. Freud (Eds.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (pp. 226-239). London: Hogarth, 1956. Vol. 12.
- Freud, S. (1915). Repression. In J. Strachey & A. Freud (Eds.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (pp. 143-158). London: Hogarth, 1956. Vol. 14.
- Freud, S. (1920). Beyond the pleasure principle. In J. Strachey & A. Freud (Eds.), *The Standard Edition of the Complete Psychological Works of Sigmund Freud* (pp. 7-66). London: Hogarth, 1956. Vol. 18.
- Freud, S. & Breuer, J. (1895). Studies on hysteria. In J. Strachey & A. Freud (Eds.), The Standard Edition of the Complete Psychological Works of Sigmund Freud (pp. 215-239). London: Hogarth, 1956, Vol. 2.
- Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. Psychological Bulletin, 84, 279-325.
- Haan, N. (1963). Proposed model of ego functioning: Coping and defense mechanisms in relationship to IQ change. Psychological Monographs, 77 (whole no. 571).
- Herrnstein, R. (1981). Self-control as response strength. In E. Szabadi & C. Lowe (Eds.), *Quantification of Steady-State Operant Behavior*. Amsterdam: Elsevier/North Holland.
- Klein, G. (1973). Two theories of one? Bulletin of the Menninger Clinic, 37, 102-132.
- Logue, A.W. (1988). Research on self-control: An integrating framework. *Behavioral and Brain Sciences*, 11, 665-709. Luborsky, L., McLellan, T., Woody, G., O'Brien, C., & Auerbach, A. (1985). Therapist success and its determinants. *Archives of General Psychiatry*, 42, 602-611.
- Messer, S. & Winokur, M. (1980). Some limits to the integration of psychoanalytic and behavior therapy. *American Psychologist*, 25, 818-827.
- Michels, R. (1983). Scientific and clinical functions of psychoanalytic theory. In A. Goldberg (Ed.), *The Future of Psychoanalysis: Essays in Honor of Heinz Kohut* (pp. 125-135). New York: International Universities.
- Pitts, F. & McClure, J. (1967). Lactate metabolism in anxiety neurosis. New England Journal of Medicine, 277, 1329-1336.
- Plato (1961) Phaedo. (trans. H. Tredennick). In E. Hamilton & H. Cairns (Eds.), *Plato, the Collected Dialogues*. Princeton, NJ: Princeton University.
- Vaillant, G. (1976). Natural history of male psychological health V. The relationship of choice of ego mechanism of defense to adult adjustment. Archives of General Psychiatry, 33, 535-545.
- Vaughan, W. & Herrnstein, R. (1985). Stability, melioration, and natural selection. In N. L. Green & J. Kagel (Eds.), Advances in Behavioral Economics, Vol. 1. Norwood, NJ: Ablex.