
COMMENTARIES

The text is not subject to copyright.
The final format is copyrighted by this journal.

Studying Self-Regulation the Hard Way

George Ainslie

*Department of Veterans Affairs Medical Center, Coatesville, PA
and Department of Psychiatry and Behavioral Science
Temple University School of Medicine*

This target article catalogs some of the most interesting available observations on failures of self-control. However, it also illustrates how the cognitive approach to self-control stays close to the ad hoc level and points to no robust principles of decision making. Reluctance to use the construct of motivation is responsible for much of this problem.

Baumeister and Heatherton use motivation haphazardly—as a cause of impulses, for instance, but not clearly as an agent in their control. Nevertheless, motivation plays a greater role here than in many articles (e.g., Perris, Blackburn, & Perris, 1988; Williams, Watts, MacLeod, & Mathews, 1988) that have followed in the wake of the “cognitive revolution” (Miller, 1988).

The realm of hypotheses permitted by the cognitive approach is summarized succinctly:

Misregulation occurs because [people] operate on the basis of false assumptions about themselves and about the world, because they try to control things that cannot be directly controlled, or because they give priority to emotions while neglecting more important and fundamental problems.

Here we see motivation stealing back into the cognitivist lexicon in the guise of its cognate, emotion. However, as an explanatory principle, this kind of motivation differs markedly from the strict maximand of the behaviorist, economist, or other utilitarian. To Baumeister and Heatherton, the motivation of emotion is only *a* reason to make a decision, and a rather disparaged reason at that, rather than *the* reason. Implied is a cognitive homunculus that evaluates emotion together with a number of other reasons for deciding and makes an autonomous choice.

Baumeister and Heatherton do depict their homunculus, the agent in “self-regulation,” as constrained by motivational weights. Unlike the dispassionate judge of

some models, this agent is reminiscent of a canoeist shooting the rapids and making calculations of forces greater than her own in order to do so. However, in this analogy there is still a decision maker who has some strength separate from that of the currents of water—the strength of her arms that she can use to steer the canoe toward what she judges to be a favorable current. The forces in the water still do not determine her steering, unless she lets them. The most parsimonious model would postulate a person whose steering direction came from the same currents that propelled the person downstream—someone who ultimately did not weigh incentives but was weighed on by them. However, models that have done so have seemed to rob the person of subtlety, to simplify her out of the free will that people usually report, and thus reduce her to a calculating machine. This has been a factor in the current disillusion with strict utilitarianism (e.g. Hollis, 1983; Schwartz, 1986).

Such simplicity exists only if we assume that the incentives that bear on a person’s behavior are discounted for delay in the kind of exponential curves that bankers use—curves that keep the same proportion between the discounted values of various goods as the delay until there are available changes. This assumption has been frequently tested over the past 20 years, and the data have regularly shown it to be false for both human participants and animals, except in some cases in which people are choosing between token rewards (Ainslie, 1975, 1992, pp. 63–80; Davison & McCarthy, 1988, pp. 96–98, 210–242; Green, Fry, & Myerson, 1994). On the contrary, these data have shown that the basic form of discounting is hyperbolic, the form described by Herrnstein’s (1961) matching law. Hyperbolic curves predict a rich, strategic interplay of motives, including the recursive processes that best fit people’s description of free will (Ainslie, 1992, pp. 200–205). I summarize Baumeister and Heatherton’s

theory of self-regulation and then argue that it contains loose ends that hyperbolic discount curves are able to tie up.

Baumeister and Heatherton's Cognitive Model

Baumeister and Heatherton do not address the question of why moment-to-moment choice does not consistently serve the person's best interests. They take the need for regulation as a given and begin by summarizing a number of social psychological experiments and clinical reports about why self-regulation succeeds or fails. They say that "certain responses are set in motion, either by innate programming, learning, habit, or motivation." For unspecified reasons, these processes sometimes prevent the person from maximizing her expected utility; however, people have a countervailing tendency to "monitor themselves" by means of "clear and consistent standards" and apply a "limited resource," dimensioned as "strength," to forestall the problematic responses. Such a response "is initiated by a combination of latent motivations and activating stimuli; self-regulation is a matter of interrupting that response and preventing it from running to its normal, typical outcome." Properties of strength are that it "should be consistent across a variety of spheres," that "when strength is depleted by demands in one sphere, self-regulatory breakdowns may occur in others," and that "it is possible to increase strength by regular exercise."

The impulses against which the strength must be applied are most likely to prevail "when attention slips off of long-range goals and high ideals and instead becomes immersed in the immediate situation." Because "emotion tends to have the effect of concentrating attention in the here and now," it is a prime culprit, although some emotions like "anticipatory guilt" may strengthen self-control.

The person must use self-regulatory strength to counteract "the short-term attentional focus caused by high emotion" by means of what Baumeister and Heatherton call *transcendence*. "Transcendence is a matter of focusing awareness beyond the immediate stimuli" so that these stimuli are seen "in the context of more distal concerns (e.g., values, goals, and motivations)." An important property of transcendence is the "asymmetry in the way many people confront internal conflicts surrounding self-regulation." That is, its continued success requires consistent choice in the direction of self-control. It collapses after "even a very brief period in which the costs [of self-regulation] seem to outweigh the benefits. The popular image in which a moment of weakness can undermine months or years of virtuous self-denial is somewhat accurate," possibly

because of additional emotional stress, "guilt or other forms of anxiety resulting from the initial indulgence." Thus, a single lapse is apt to lead to a binge, as in Marlatt and Gordon's (1980) abstinence violation effect.

In addition to the process just described, called *underregulation*, self-control may suffer from "misregulation" caused by "(a) misunderstood contingencies, (b) quixotic efforts to control the uncontrollable, and (c) giving too much priority to affect regulation."

Baumeister and Heatherton often allude to a marketplace model in which alternatives compete on the basis of some common dimension—strength, height (of emotion), cost, and even reward—but they shrink from concluding that this marketplace strictly determines choice. Without such a unifying principle, the crucial determinants of choice get attributed to unmotivated processes that interact with the motivational (or economic) factor in unspecified ways. Among the more important questions not addressed are:

- What constrains the person's establishment of standards and what determines how "clear and consistent" they have to be?
- How does self-regulatory strength get differentiated from impulsive strength and why does the timing of the contest affect its outcome so greatly?
- What is added to the mere motivation to achieve long-range goals that converts it into strength, with the enumerated properties of generality, exhaustion, and growth with practice that do not characterize motivation per se?
- Is there more to transcendence than the simple motive to obtain long-range goals?
- Why should the motives of the "immediate situation" be asymmetrical with the motives for "long-range goals and high ideals"?
- Given this asymmetry, what constrains the person's possible strategic responses to it, for instance when deciding whether or not to "catastrophize the initial lapse as a way of preventing it"?

I expected an even more significant question to be covered under the "Misregulation" section but it was not: Given the great deal of clinical symptomatology that seems to involve overregulation (e.g., obsessive-compulsive character, compulsive dieting, emotional constriction, and other behavioral rigidities), is it wise to assume that self-regulation operates entirely in the person's long-range interest? Perhaps regulation is an imperfect solution to the problem of "[immersion] in the immediate situation"; but without a theory of why that immersion occurs to begin with, such a possibility is obviously hard to discuss.

A Behavioral Economic Alternative

A hyperbolic discount function has seemed counter-intuitive to many people, perhaps because it predicts the very problems of self-regulation that are described in the Baumeister and Heatherton target article, and requires something very like the principles described there to prevent seriously maladaptive behavior. Indeed, these authors' extensive review of the self-control literature forms a detailed survey of the implications of hyperbolic discounting, saving only the unifying mechanism of this discount process itself.

First of all, hyperbolic discount curves provide a fundamental mechanism for impulsiveness. Because hyperbolic curves do not stay proportional to each other as exponential curves do, but disproportionately value imminent events, they often predict a temporary preference for the smaller but earlier pair of alternative goods.¹ These curves put a person into a limited-war relationship with the person's future selves: When facing an overly influential temptation, the self that views it from a distance will prefer to avoid it and the self that gets close will prefer to obtain it over its better long-run alternative; both selves will prefer that subsequent selves pick the better, later alternative when similar choices recur in the future. This mechanism will regularly produce impulsiveness without a need for special exceptions to the motivational market like "latent motivations" and "activating stimuli." Indeed, I have argued elsewhere that the commonly cited pathogens of impulsiveness—conditioned emotions, repressed motives, superstitious learning, and so forth—do not represent separate processes of motivation and would not be expected to distort the motivational impact of the person's prospects (Ainslie, 1992, chap. 2).

Hyperbolic discount curves also answer the question of why a person needs standards, as opposed to mere ordered preferences, and what the efficacy of these standards depends on. A single larger and later reward may often be outweighed by a smaller, imminent one; however, a series of such rewards is less apt to be outweighed by a series of smaller and earlier ones, because only the first in the series of smaller alternatives is disproportionately valued. Insofar as a person perceives a series of choices to be similar, the prospects that are at stake in each choice are not just those of the alternatives being literally chosen: If the person sees

herself choose the larger and later reward this time, she increases her expectation of choosing it on subsequent occasions; if the person sees herself choose the smaller and earlier one, she will have less reason to believe she will do anything different in the future. A whole set of rewards has come to be staked on each choice. In other words, a "standard" operates like a diet—If I choose the cookie now before me I can expect to suffer more than its trivial caloric load; I will reduce my faith in my diet. Thus, personal standards (my own term is *personal rules*) are self-enforcing in the same way as are contracts between people who expect to deal with each other repeatedly (Klein & Leffler, 1981) or tacit agreements to cooperate in a repeated prisoner's dilemma (Axelrod, 1984).

This model of standards, which will actually serve as a model of willpower, creates a very specific hypothesis about "strength." The strength that defends a given standard is the aggregate prospective reward that is at stake if that standard should fail. Such strength will indeed accumulate with practice, as Baumeister and Heatherton predict in their target article; however, it will not fatigue in any regular way except through being overwhelmed by the sheer size of a temptation, and its generality will depend on what choices are perceived as precedents for what standards. Indeed, the greatest threat to people's standards seems to come not from an overwhelming temptation but from their attempts to claim a current choice as an exception to their standards. The effort of attention hypothesized in the target article would better be seen as a contest for interpretation: If a situation is rare enough a person can make it an occasion to gratify an urge without damaging prospects for self-control in the future; however, the urge itself may make the person reckless in claiming just-this-once rationales (rationalizations).

In this internal bargaining model, strength of will is a recursive process like the stock market, where the market's present will to buy depends to a great extent on people's forecasts of its future will to buy. As in the stock market, a significant fall in trust can lead to a cascade of selling, causing the market to fall much faster than it ever rises. Here is a mechanism for the striking abstinence-violation effect that Baumeister and Heatherton mention—in the case of alcoholism so striking as to often be attributed to a physical reflex (shown not physical in Maisto, Lauerman, & Adesso, 1977). However, in a motivational model, the "stress" of a lapse is a fall in the prospect for sobriety, and it is the very threat of this happening that forms the strength of the will. It may even be that the emotion of guilt is composed largely of the sense of damaged self-control, rather than being an independent contributor to this damage.

¹Space does not permit a mathematical exposition. The best fit of available discounting data is afforded by the generalized form of Mazur's equation, $\text{Value} = (\text{Value if immediate}) / (\text{Constant}_1 + [\text{Constant}_2 \times \text{Delay}])$ but all hyperbolic curves will have similar properties (Ainslie, 1992, pp. 60–76; Mazur, 1987).

This view of strength supplies an explicit hypothesis about the constraints on such ostensibly cognitive tasks as the setting of standards, the achievement of transcendence, and the choice of whether to have "zero-tolerance beliefs." Cognitive therapists are in the habit of assaulting zero-tolerance beliefs as irrational; however, to put up nothing important as hostage for a personal rule would be to rob that rule of all effectiveness. Minimizing damage from the abstinence violation effect is a delicate task, for it involves finding rationales by which the person can put enough prospects at stake to deter lapses but not enough to cripple recovery if a lapse occurs. Baumeister and Heatherton's analogy of a battle commander deciding how much to commit reserves is apt.

In the model I have described, the will is not a normalizing force that erases aberrations of preference. Rather, it is an imperfect committing device that a person cobbles together to a greater or lesser extent from reward-predicting processes that doubtless evolved in lower species for other purposes. The aggregation of series of options into interdependent choices can allow people to discount the future as if their discount curves were nearly exponential (Ainslie, 1991), but this expedient differs in many ways from an exponential preference pattern that would occur spontaneously. The person pays for increased impulse control with several motivational distortions (Ainslie, in press):

1. A legalistic style in which events are evaluated more as precedents than as experiences in their own right.
2. Dominance of moderately long-range, highly quantifiable options over still longer range but subtler ones, as in the victory of the compulsive dieting of anorexia nervosa over richer activities like human relationships.
3. Intense motivation to abandon efforts of will that have led to lapses, lest the will lose credibility generally, but at the cost of entrenching symptoms like addiction, panic, or pathological grief.
4. Distorted information processing so as to avoid the perception of having violated a rule, leading to the symptoms often attributed to repression and dissociation.

I would have called these developments the major forms of misregulation, rather than the three cases that Baumeister and Heatherton list. The first two of Baumeister and Heatherton's cases seem to be another form of describing underregulation—insufficient strength for the chosen objective; the last sounds like an example of rationalization, a natural consequence of the legalism that goes with reliance on willpower.

A Final Perspective

Not so long ago the behaviorists and the psychoanalysts fought bitterly about who had the proper science of motivational conflict. After a flirtation with experimental science, the analysts turned away from it; the behaviorists were seduced into the philosophical controversies about mentalism that wound up defining their identity to the rest of the scientific community. Ironically, these two schools shared the fundamental discipline that all choice must be accounted for with strict economic principles, exactly the discipline lacking in the cognitive approach that has largely supplanted them both.

Baumeister and Heatherton review the problem of self-regulation skillfully. It is the very fact that they capture a great deal of its complexity that reveals the theoretical inadequacy of the tools they use. With their various motivational concepts they seem to be reaching out to recapture the economic insights of behaviorism, whose major contribution to science was not its logical positivism but the parametric motivational experiments that produced the matching law. I hope they will feel encouraged to explore further in this direction.

Notes

I thank Barbara Gault for extensive comments.

George Ainslie, Department of Veterans Affairs Medical Center, Coatesville, PA 19329.

References

- Ainslie, G. (1975). Specious reward: A behavioral theory of impulsiveness and impulse control. *Psychological Bulletin*, 82, 463–496.
- Ainslie, G. (1991). Derivation of "rational" economic behavior from hyperbolic discount curves. *American Economic Review*, 81, 334–340.
- Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person*. Cambridge, England: Cambridge University Press.
- Ainslie, G. (in press). The dangers of willpower. In J. Elster & O.-J. Skog (Eds.), *Getting hooked: Rationality and the addictions*.
- Axelrod, R. M. (1984). *The evolution of cooperation*. New York: Basic.
- Davison, M., & McCarthy, D. (1988). *The matching law: A research review*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Green, L., Fry, A., & Myerson, J. (1994). Discounting of delayed rewards: A life-span comparison. *Psychonomic Science*, 5, 33–36.
- Hermstein, R. (1961). Relative and absolute strengths of response as a function of frequency of reinforcement. *Journal of the Experimental Analysis of Behavior*, 4, 267–272.
- Hollis, M. (1983). Rational preferences. *The Philosophical Forum*, 14, 246–262.
- Klein, B., & Leffler, K. B. (1981). The role of market forces in assuring contractual performance. *Journal of Political Economy*, 89, 615–640.
- Maisto, S., Lauerma, R., & Adesso, V. (1977). A comparison of two experimental studies of the role of cognitive factors in alcoholics'

- drinking. *Journal of Studies on Alcohol*, 38, 145-149.
- Marlatt, G. A., & Gordon, J. R. (1980). Determinants of relapse: Implications for the maintenance of behavior change. In P. O. Davidson & S. M. Davidson (Eds.), *Behavioral medicine: Changing health lifestyles* (pp. 410-452). Elmsford, NY: Pergamon.
- Mazur, J. E. (1987). An adjusting procedure for studying delayed reinforcement. 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* (pp. 55-73). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Miller, L. (1988). Behaviorism and the new science of cognition. *The Psychological Record*, 38, 3-18.
- Perris, C., Blackburn, I. M., & Perris, H. (Eds.). (1988). *Cognitive therapy: Theory and practice*. London: Springer-Verlag.
- Schwartz, B. (1986). *The battle for human nature: Science, morality and modern life*. New York: Norton.
- Williams, J. M. G., Watts, F., MacLeod, C., & Mathews, A. (1988). *Cognitive psychology and emotional disorders*. New York: Wiley.

Failures in Self-Regulation: Energy Depletion or Selective Disengagement?

Albert Bandura

Department of Psychology
Stanford University

Baumeister and Heatherton address the all too common social transgressions, dysfunctional adaptations, and inhumanities in terms of widespread breakdowns in self-regulation. The issues they raise are of considerable import both theoretically and socially. This commentary examines the nature of their conceptual scheme and contrasts it with a more ecologically oriented sociocognitive theory of self-regulation.

Baumeister and Heatherton's conception of self-regulatory failure is grounded in the negative feedback-loop model. In this psychocybernetic system, a perceived negative discrepancy between a sensed feedback and an inner referent triggers adjustments to reduce the negative incongruity. There is a serious question whether the feedback loop is equipped to bear the explanatory burden placed on it given the complexities of human self-regulation. The occurrence of perceived negative discrepancies, in fact, says little about how people will react to them. Some assuredly generate better strategies and redouble their efforts to match their internal standard, others lower their standard and become resigned to a humbler aspiration, still others continue to impose on themselves the elusive standard but debilitate their efforts through gnawing despondency, and some, surprisingly, even raise their standard in the face of failure. The explanatory challenge is why the same level of perceived negative discrepancy produces such variable cognitive, behavioral, affective, and motivational effects. Three self-reactive factors in social cognitive theory (Bandura, 1991a; Bandura & Cervone, 1986) do a pretty good job in predicting whether perceived negative discrepancies will be motivating, demotivating, depressing, or uplifting. These factors include perceived self-efficacy to fulfill given standards, affective self-reaction to substandard perfor-

mance, and readjustment of personal standards in light of the progress being made.

To capture the complexities of human self-regulation requires a control system with evaluative agentive properties known to govern self-directedness. These include (a) proactive adoption of standards rooted in a value system and subserving advantageous purposes; (b) self-appraisal of personal efficacy to fulfill given standards; (c) anticipatory regulation of the strategies and effort needed to turn cognized standards into reality; (d) outcome expectations for fulfillment or unattainment of the standards; (e) affective self-evaluative reactions to one's performances; and (f) metacognitive activity concerning the accuracy of one's efficacy appraisals, the suitability of one's standard setting, and the adequacy of one's effort and strategies.

The feedback-loop model has recently come under heavy fire. Locke (1991, 1994) documented how adherents to this version of control theory have now grafted so many ideas from other theories on the negative feedback loop to remedy its prediction problems that control theory has lost its distinctiveness.

Discrepancy reduction clearly plays a role in any system of self-regulation. However, self-regulation via negative discrepancy tells only half the story and not necessarily the more interesting half. People are proactive, aspiring organisms. They set for themselves performance challenges that create motivating discrepancies to be mastered. Self-regulation thus involves a hierarchical dual-control process of disequilibrating discrepancy production followed by equilibrating discrepancy reduction. It requires proactive control as well as reactive control. In some circles the challenges people set for themselves may serve antisocial or evil purposes. It takes proficient self-reg

