APA 7/16/02

The Effect of Hyperbolic Discounting on Personal Choices

George Ainslie

Veterans Affairs Medical Center, Coatesville, PA
and
Temple University Medical College

151 Veterans Affairs Medical Center
Coatesville, PA 19320
610 383-0260

Keynote speech to the thematic session, "Personal Choice and Change"
Presented at the annual convention of the American Psychological Association
at Chicago, IL, August 22, 2002 at 1:00 PM.

Abstract

Behavioral science has had trouble accounting for addictions and other knowingly self-defeating behaviors. It generally depicts them as pathological changes in an innate, natural rationality. However, there is now ample experimental evidence that all behaving organisms have a basic tendency to devalue expected rewards as a hyperbolic function of delay, which is much more deeply bowed than the conventional exponential function. This finding implies that organisms will often form temporary preferences for smaller-sooner (SS) rewards over larger-later (LL) ones when the SS rewards are imminent, and thus are innately impulsive. It also implies a motive for farsighted organisms like humans to avoid these temporary preferences. The most versatile impulse-avoidance tactic is willpower, for which I propose a two part mechanism: (1) Bundling choices into whole categories increases the influence of the LL alternatives, a phenomenon predicted by hyperbolic but not exponential curves; it has actually been observed in recent human and animal experiments that I describe. (2) Perceiving your current choice as a precedent predicting future choices forms *de facto* bundles, an effect for which thought experiments provide the best evidence.

Utility theorists have also been at pains recently to account for why people invest importance in other people's experiences, particularly "the problem of altruism." Here, too, the phenomenon of hyperbolic devaluation of expected rewards offers a solution. Emotional experience is the most important source of reward in societies whose material needs are highly satiated. Emotion is within a person's power to generate, but generating it at will leads people to harvest its rewards prematurely because of a hyperbolically based impatience for SS reward. Thus only that emotionality which is cued by unpredictable events will escape rapid inanition into a daydream. This process can be expected to create an ongoing need for surprise, of which the kind occasioned by other people will be the most salient. The result is that vicarious experience becomes a primary good, rather than the byproduct of an interpersonal game strategy or an internalization of norms.

Thus a single underlying phenomenon suggests a theoretical integration of impulsiveness, willpower, and subtler processes like altruism, which have been anomalies for conventional motivational theory. Conventional utility theory describes only those choices that are made under conditions favorable to the exercise of willpower, and thus represents a special case within a more present-focused motivational universe.

Keywords

Hyperbolic discounting, will, impulsiveness, self-control, addictions, empathy, altruism, utility theory

Self-defeating behavior has been accounted the greatest preventable cause of death in the modern world. Yet people continue to smoke, drink, take drugs, overeat, and indulge in unsafe sex. In addition, the misery caused by pathological gambling, credit card abuse, and an array of other habits down to sheer procrastination is rarely fatal, but still hard to explain in people who are fully aware of the consequences. Behavioral science has done no better than folk psychology in accounting for these behaviors. I assert that its error has been to look for pathological processes that attack a basic, natural rationality, while in fact people share with nonhuman animals an elementary trait that makes us irrational, at least for living in the developed world. A realm of rational utility-maximizing exists, of course, and is well described by classical economics; but this realm is unnatural, in the sense that it requires continual effort to maintain it on top of a shifting motivational base. What has been called "rational choice theory" (Korobkin & Ulen, 2000)[1] describes a special case within a much less rational motivational system.

I base this sweeping conclusion on two kinds of evidence: parametric experiments on both human and animal subjects, which have unequivocally found a discount curve for delayed rewards different from the one assumed by conventional utility theory; and the efficiency of this discount curve in accounting for higher order phenomena of human choice with parsimonious assumptions. Hypotheses derived from this curve are only beginning to be tested by controlled experiment, and many of them may never be testable in this way, because they involve recursive processes. However, I argue that *thought experiments* as developed by the philosophy of mind can also be reliable tests of their validity, and that these support the hypotheses I will present. At the very least, this approach has generated the first fully reductionistic theory of willpower and several other high-order phenomena, among which I will discuss briefly vicarious reward and altruism. It thus represents a means of integrating some of the diverse phenomena that have been described as anomalies for conventional utility theory.

Conventional utility theory has long acknowledged a discount curve for the value of expected events as a function of delay. However, this curve is simply exponential, the subtraction of a constant proportion of remaining value for each unit of delay. This curve does not depend on when the event is expected-- only on knowledge of its value at some other specific time and the rate of discount. Thus at a discount rate of ten percent the expected value of $900 delayed for a year is the same as the value of $1000 delayed for two years ($810 in both cases), and the values of these two amounts will remain indistinguishable when any further delay is added to both ($729 another year earlier, and so on). The exponential is the only form of curve that will not cause shifting preferences among events due at different moments, purely as a function of elapsing time. Exponential curves are esthetically satisfying, not only because they are mathematically

[1] Rational choice theory seems to be what non-economists call the broader implications of classical economics, one of the cornerstones of which is the maximization of utility. Those behavioral sciences that deal with motivation have all taken it as at least the norm of rationality, if not (as economics seems to) an actual constraint on choice.

tractable but also because of the consistency of behavior they imply. These properties have led people to take them as basic, as the obviously rational way to evaluate future events. Nevertheless, exponential curves are only normative-- They do not necessarily describe people's actual valuations. The prevalent intuition that people naturally discount the future exponentially, and deviate from this only when there is some abnormality, is not necessarily correct.

Hyperbolic Discounting

The overwhelming preponderance of experimental evidence indicates that this intuition is indeed not correct. In one way or another it is possible to ask subjects as diverse as rats, pigeons, retarded people and economics students how much they would value the prospect of a particular amount of a good-- food, rewarding brain stimulation, money, access to a game, relief from noxious noise-- at various delays. Except in the case where large or repeated amounts of money are at stake, their answers always describe a discount curve that is more deeply bowed than an exponential curve (Figure 1; Green, Fry, & Myerson, 1994; Kirby, 1997; Mazur, 2001).

1.

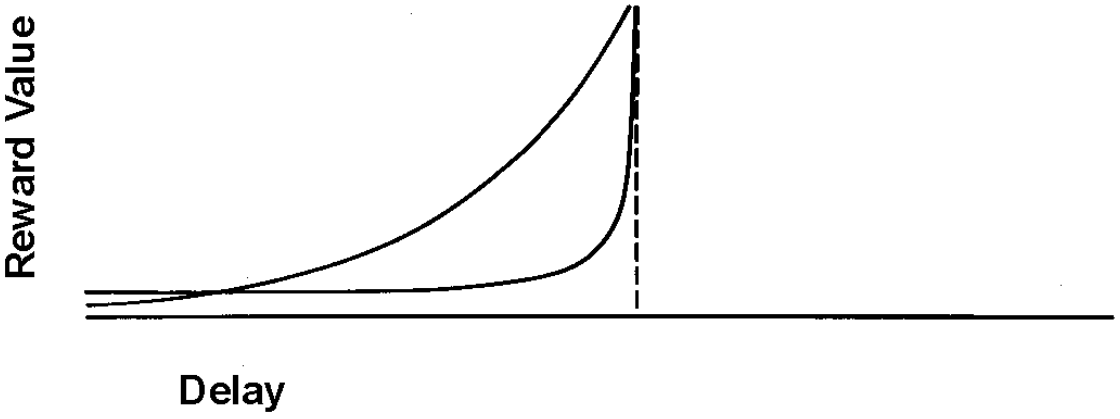

Figure 1. An exponential discount curve and a hyperbolic (more bowed) curve from the same reward. As time passes (rightward along the horizontal axis), the motivational impact-- the value-- of the goal gets closer to its undiscounted size, which is depicted by the vertical line.

The data are best fitted by an inverse of [delay multiplied by a factor describing individual steepness of discounting], with a small constant added to the denominator to

reflect the fact that values do not approach infinity as delays approach zero (Mazur, 1987):

$$\text{Value} = \frac{\text{Value if immediate}}{\text{Constant (1)} + (\text{Delay} \times \text{Constant (2)})}$$

The trouble with this formula for practical calculation is that re-evaluation with the passage of time requires a fresh computation from the current moment to the moment of reward at every point. To avoid this nuisance, Laibson has suggested making exponential curves more hyperboloid by inserting an additive term in the conventional exponential formula (1997). The result is a step function that describes overvaluation of imminent events but leaves the discounting of later events exponential, a solution also proposed by Simon (1995). This kind of curve does not fit the experimental data as well as a hyperbolic curve, but it nevertheless predicts some observed anomalies of human economic choice, such as a preference for illiquid savings-- those that cannot be tapped at will without a penalty (Harris & Laibson, 2001). However, the great importance that hyperboloid discounting has for motivational theory does not arise from the exact shape of the discount curve, but rather from the implication that any shape more bowed than an exponential curve has: that preference among rewards due at different moments will tend to change as a function of elapsing time (Figure 2).

2.

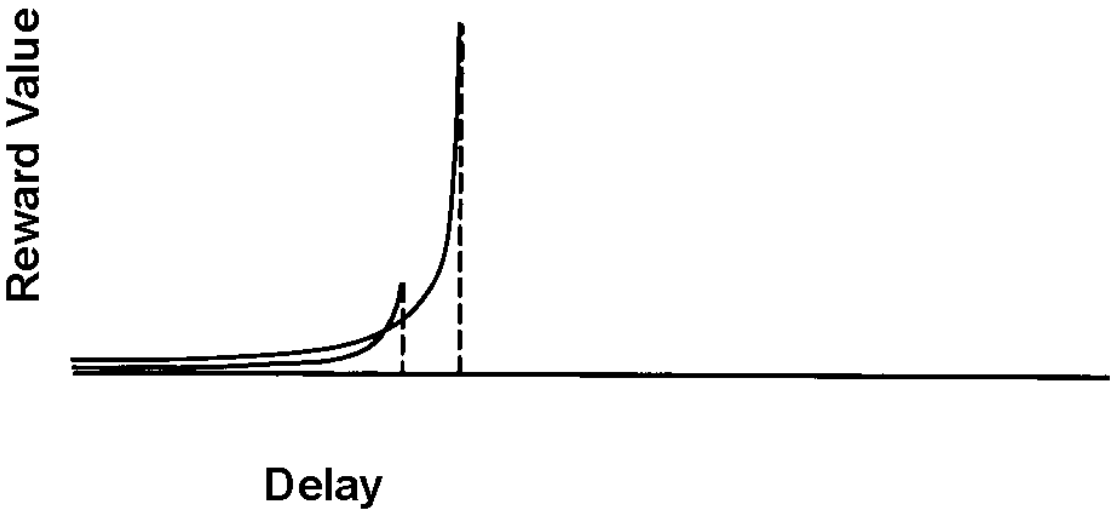

Figure 2. Hyperbolic discount curves from two rewards of different sizes available at different times. The smaller-sooner reward is temporarily preferred for a period before it is available, as shown by the portion of its curve that projects above that from the larger-later reward.

Deeply bowed discount curves predict that people will regularly form temporary preferences and thus that the self at any given moment will be in a relationship of limited warfare (Schelling, 1960, pp. 53-80) with expectable future selves. That is, it will share with them the preferences that are not affected by imminent reward, but not those that are. The conflicting motives do not come to equilibrium, because they are dominant at different times. In a bare bones model of utility maximization based on hyperbolic discounting, all rewards that actually happen support the learning of activities to get them, just as habitats in nature select for species that can exploit them. With experience many of these activities come to be strategic, means of forestalling incompatible goals that are dominant at other times. Those activities selected by a particular reward could be called its *interest*, just as economic and political interests can be identified on the basis of their objectives. This model has a radical implication for personality theory: that the person behaves as a unit only insofar as her longest range interest, the one that is stable, has adequate influence over her shorter range interests. Fundamentally a person is a population of interests. With allowances for the fact that these interests can exert influence only sequentially, not simultaneously, dealing with this population should be much like dealing with a population of individuals.

At this point intuition begins to object. Temptation may well follow a deeply bowed curve, and addicts may indeed express preferences for future goods, both addictive and nonaddictive, in curves that are not only hyperbolic but steeper than other peoples' (Vuchinich & Simpson, 1998; Bickel et.al., 1999). But many people learn to function effectively in situations where a hyperbolic pattern of preferences would pump money into competitors' pockets; and the self, in the sense of ego, usually means an organ that is not just momentary but integrated over time. Hyperbolic discounting may describe passion, the objection continues, but the faculty of reason has been recognized since Plato's time to be capable of governing passion.

However, intuition is formed at the experiential level, and is not necessarily in contact with the mechanism of either temptation or self-control. The job of a theory is to take basic processes that have been observed empirically and build processes out of them that wind up matching our intuitions. Existing theories have not done this. The currently dominant psychological approach to choice, cognitive theory, basically expresses intuition in abstract language, without making any hypotheses about how reason might be governed by incentives, and even if not *governed*, how *influenced* by incentives together with some other factors in tandem. (A typical formulation: "When all higher-level control is removed temporarily and sequence control is given free rein, people... are more responsive to passing cues of the moment that touch off sequences of action."-- Carver & Scheier, 1990, p. 21.) Classical utility theory, of course, views the choice-maker as a straightforward estimator of the amounts, probabilities, and delays of environmental events, with no provision for temptation or self-control. Hyperbolic discounting theory, having shifted the central question from how passions arise to how rationality arises, must try to satisfy intuition with a mechanism by which conflicting momentary selves

can generate a somewhat consistent self that partially resembles Economic Man, but without entirely losing the notorious human tendency to be irrational.

Strategic Self-Commitment

A basic tendency to devalue the future hyperbolically accords with observations of the intractability of addictions and other self-destructive behaviors, but raises the question of how they are controlled. Fortunately, experimental research has found not only the predicted temporary preference for smaller, sooner (SS) over larger later (LL) rewards when the SS rewards become imminently available, but also several ways that people and even nonhuman animals learn to constrain forestall a future choice of SS rewards when the LL are dominant. In the simplest paradigm, pigeons that are given choices between a SS food reward for pecking a key and a LL food reward for not pecking often learn to peck a key presented earlier, the only effect of which is to prevent the key that produces the SS reward from becoming available (Ainslie, 1974). This finding suggests that familiar human devices like the Antabuse that makes alcohol sickening and savings schemes that charge a premium for *reducing* liquidity are not peculiarities of our culture, but depend on the basic configuration of discounted prospective reward. Over the last three decades Walter Mischel and his co-workers have done a number of experiments with 4 to 6 year old children, showing that in this age range they learn to use both distraction of their attention and control of their emotions ("cool thoughts") to wait for an LL food reward in the face of temptation by a SS one (e.g. Metcalfe & Mischel, 1999). An economist, Juan Carrillo, has likewise described the "value of ignorance (in the form of not acquiring [even] free information)" in avoiding financial temptations (1999).

These observations represent three strategies of influencing future behavior: changing the environment, controlling attention, and cultivating mental processes that have some momentum, like emotions (reviewed more fully in Ainslie, 1992, pp. 125-142, and Ainslie, 2001, pp. 74-78). Perhaps the most important example over the years has been making yourself susceptible to influence by other people, which is clearly the most persuasive environmental factor. A hothead may benefit from her friends' being calmer, and an overeater find motivation in social competition to stay slim. To some extent social forces can exert pressure against impulses, as when a competitive market raises the stakes for being careful and consistent with money. However, this strategy is vulnerable to impulses that strike everyone together, leading to "the madness of crowds," and would be actually counterproductive against impulses that let others exploit the person-- an urge to buy friendship, for instance. As the urbanization of society increases the number of people with whom an individual is in contact, the dangers of being vulnerable to influence increase, which is probably a factor in the increased emotional guardedness observed in cosmopolitan societies (e.g. Stearns, 1994).

Will as bundling of choices. Even together the three self-control strategies just described are unlikely to be enough to account for what we sense to be rational, or even consistent-- for the coherence of self that clinicians call ego strength. Common speech calls this capacity will or willpower. Will was a central topic for the first behavioral scientists in the nineteenth century, but eluded study by controlled research, and was

largely abandoned in the twentieth century. However, philosophers of mind still discuss the question of how we achieve consistent volition or "resolute choice." They say that following a plan is rational *per se*, perhaps because the a person has "a sense of commitment to a plan initiated by [a prior] self (McClennen, 1990, pp. 157-161)," or a "concern with how she will see her present decision at plan's end (Bratman, 1999, pp. 50-56)." The most robust attribute, which has been hypothesized since Plato, is that consistency depends on how much an intention is based on principle rather than just the incentives present in a particular situation (reviewed in Ainslie, 2001, pp. 78-81). These attributes are intuitions rather than parts of a coherent mechanism. However, the properties they suggest-- mental commitment, concern with a future self's perception, and, especially, membership in larger categories-- delineate what the underlying mechanism is apt to include.

The hyperbolic shape of the discount curve supports the suggestion that deciding according to principle is the key to such a mechanism. Hyperbolic discount curves decline rapidly over short delays, but at long delays they decline slower than all but the shallowest exponential curves. This gives them a property that exponential curves do not have: that adding the curves from several choices together increases the summed discounted value of LL rewards relative to SS rewards. This is because only the nearest SS rewards are overvalued (Figure 3).

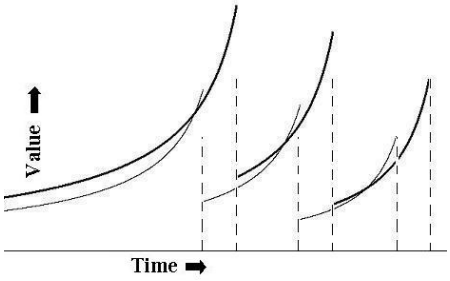

Figure 3a

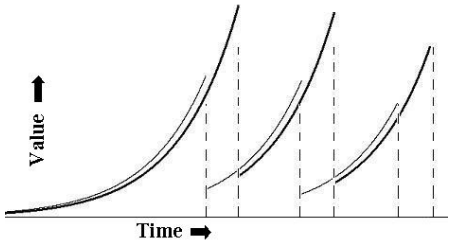

Figure 3b

Figure 3a. Summed *hyperbolic* curves from a series of larger-later rewards and a series of smaller-sooner rewards. The dashed vertical lines represent the value of the reward when immediate, and each discount curve represents the discounted value of that alternative when summed with all other like rewards occurring later in time (to the right). At the beginning of the series, preference for the series of larger rewards is consistent By contrast, the curves from just the final pair of rewards indicate a period of temporary preference for the smaller-sooner reward when it is imminent.

Figure 3b. Summed *exponential* curves from the two series of rewards shown in Figure 1a. Again, the dashed vertical lines represent the value of the reward when immediate, and each discount curve represents the discounted value of that alternative when summed with all other like rewards occurring later in time (to the right). Summing does not change the relative heights of the curves.

This property is testable by controlled experiment, and is in fact found: Students who are given choices of SS vs. LL sums of money or SS vs. LL slices of pizza at weekly intervals choose the LL more if they choose for five weeks all at once than if they are given the choices one by one (Kirby & Guastello, 2001). An experiment with students may always be suspected of depending on something in the way it is presented to them that gives away the experimenters' hypothesis-- and, by implication, wish-- or at least on something in the subjects' acculturation that moves their choice beyond where the ostensible rewards would move it by themselves. However, the bundling effect can also be seen in nonhuman animals. If rats are given choices of SS vs. LL sugar water at 6 second intervals, they prefer the LL more if they choose the next three deliveries at once than if they choose singly (Ainslie & Monterosso, in press), showing again that hyperbolic discount curves have a direct influence independent of culture.

Bundling from intertemporal bargaining. Still, people do not usually have the opportunity to commit whole series of their future choices at once. Furthermore, the mechanism of willpower must be intrapsychic, and remain responsive to changes of circumstance. To fit the experience of willing, the intending self must influence future selves without entirely controlling them, and must do so without gimmicks. To be of any theoretical use, it must also do so without resorting to a faculty that itself possesses management power, an ego or homunculus that buffers the self from strict determination. The hyperbolic shape of the discount curve dictates that the person in her present motivational state must find ways to affect future motivational states, specifically to appeal to the interest she has in common with them despite the limited war she is necessarily conducting with them. Every night she wants to drink too much but also not to be a drunk generally, or overeat without becoming fat, or overspend without becoming a spendthrift, or duck out of an impending speaking engagement without losing the courage to speak in public generally. The question that remains to be answered is whether there is a mechanism for her to make these choices all at once, without appealing to a special organ of rationality that is somehow wired to override the relevant motives.

The answer is intuitively obvious, but has lacked a rationale in conventional utility theory: If you have an alcoholic tendency, your reason to refrain from getting drunk is to maintain your belief that you *can* refrain from getting drunk, because the greatest part of your expectable reward depends on this belief; if you are worried about your weight, your reason to stick to your diet on any particular occasion is to maintain the credibility of this diet. You have to ensure the cooperation of your own future selves in the same way that you ensure the cooperation of your opponents in a limited warfare situation: You recognize the situation as a repeated game resembling the prisoner's dilemma, and choose in light of the knowledge that each choice is important as a precedent, a test case, in addition to whatever intrinsic value it has. This is just the logic of the "self-enforcing contract" (Klein & Leffler, 1981) applied to successive selves. The more a given choice becomes a test case, the more your expectation of future preferences and thus of future reward depends on your current choice. In a recursive situation, where a lapse creates an expectation of further lapses, a person does to a large extent choose whole series of rewards at once. The experience gets called resolve or intentionality, and the principle that serves as a criterion for cooperation gets called a *personal rule*.

Described as a game, this process sounds artificial. Certainly people's recognition of a prisoner's dilemma in ordinary volition is tacit at best. However, there are many familiar situations in which we monitor our current performance to predict a later outcome, often despairing suddenly and decisively if our prediction falls below a threshold of confidence. J. M. Russell describes seasickness as an example:

> I suspect that I may be getting seasick so I follow someone's advice to "keep your eyes on the horizon"... The effort to look at the horizon will fail if it amounts to a token made in a spirit of desperation... I must look at it in the way one would for reasons other than those of getting over nausea... not with the despair of "I must look at the horizon or else I shall be sick!" To become well I must pretend I am well (1978, pp. 27- 28).

Darwin said that emotions in general follow this pattern:

> The free expression by outward signs of an emotion intensifies it. On the other hand, the repression, as far as this is possible, of all outward signs softens our emotions. He who gives way to violent gestures will increase his rage; he who does not control the signs of fear will experience fear in greater degree (1872/1979, p. 366).

Anxiously hovering over your own performance is commonly noticed in behaviors that are recognized to be only marginally under voluntary control: summoning the courage to perform in public (versus what comedians call "flopsweat") or face the enemy in battle, recall an elusive memory, sustain a penile erection, or, for men with enlarged prostates, void their bladders. To seem to be succeeding increases the actual likelihood of success.

As with interpersonal bargaining, what is possible in intertemporal bargaining depends on what markers are available to distinguish cooperation from defection and from choices that are not relevant. The odds that an alcoholic will achieve sobriety are much greater than that an overeater will achieve normal weight, not because food is a stronger reward than alcohol but because there is a unique boundary, or *bright line*, between some drinking and no drinking at all, but many possible lines between eating too much and eating the right amount.

Evidence from thought experiments. When the behavior of a group depends on this kind of recursive self-examination, it may undergo sudden shifts, as when there is a bubble or panic in the stock market; or it may be fortified by the process, sometimes generating what is significantly referred to as "the national will." My hypothesis is that the temporary preference phenomenon which hyperbolic curves produce represents an intertemporal version of the same tacit bargaining process. However, it will be hard to test the intertemporal bargaining hypothesis by controlled research, beyond the studies that established the effect of bundling choices together. Analog studies of interpersonal bargaining can lend support (e.g. Monterosso et.al., 2002), but an experimenter cannot manipulate the factors within an individual's recursive self-perception process. To clarify the components of intertemporal bargaining I advocate a technique that is

heterodox in psychology but a mainstay of the philosophy of mind, the thought experiment (Sorensen, 1992). Not at all like the century-old exercises that discredited introspectionism (Boring, 1950), thought experiments are more akin to the techniques that linguists use to elicit the grammar of native speakers. Here is one proposed by John Monterosso:

Consider a smoker who is trying to quit, but who craves a cigarette. Suppose that an angel whispers to her that, regardless of whether or not she smokes the desired cigarette, she is destined to smoke a pack a day from tomorrow on. Given this certainty, she would have no incentive to turn down the cigarette— the effort would seem pointless. What if the angel whispers instead that she is destined never to smoke again after today, regardless of her current choice? Here, too, there seems to be little incentive to turn down the cigarette—it would be harmless. Fixing future smoking choices in either direction evidently makes smoking the dominant current choice. Only if future smoking is in doubt does a current abstention seem worth the effort. But the importance of her current choice cannot come from any physical consequences for future choices-- hence the conclusion that it matters as a precedent (Monterosso & Ainslie, 1999).

"Kavka's problem" comes at the same question from a different angle (Kavka, 1983): Suppose an eccentric billionaire offers you a million dollars if you will *intend* to drink a harmless but exceedingly noxious toxin. The intention can be documented by a brain scan, and will earn you the million whether or not you subsequently drink the toxin. Kavka's question is whether, after intending to drink, it is rational for you to go through with it; and whether, if you think in advance that it would not be rational, you would be able to muster the intention. The tease of this problem has been that, although it feels to most people that somehow they should go through with the drink, it is impossible to specify why in terms of rational choice theory. I assert that this quandary demonstrates the need for intertemporal bargaining theory, which readily supplies the missing piece: The mechanism of intention is the person's belief that she will do the intended thing if possible. To intend and be foresworn damages the faculty of intention. If and only if fulfilling your intention is important as a precedent, it is rational to carry out an intention that you formed knowing that fulfilling it would not be literally necessary. (For a fuller presentation and more examples, see Ainslie, 2001, pp. 125-140.)

Limitations of Intertemporal Cooperation. Cooperation in a repeated prisoner's dilemma will be the most stable solution to the ongoing conflict of successive motivational states. Self-control by this method depends on the person's ability to specify categories of choice with criteria clear enough for her to know whether or not each choice is a cooperation. The commonest example of such a criterion is arguably the use of an exponential formula for discounting those future goals that can be measured in cash. That is, if you make a personal rule to choose among these goals on the basis of the currently "rational" financial rate you not only become consistent in achieving your long range goals, but also gain an advantage over competitors who have not achieved as much consistency. But even these advantages do not seem usually to be enough motive to sustain a rule to make all spending choices according to this discount rate: People put their money into different "mental accounts" that vary in how readily they allow

themselves to access it for current wants (Shefrin & Thaler, 1988). It is within the most protected account, the one they use for capital investments, that their choices are apt to be governed by the laws of classical economics. People leave themselves money to spend spontaneously just to avoid confronting a strict rule with an immediate urge, since a lapse impairs their intertemporal cooperation. Not only will the contents of spending money or petty cash accounts be disposed of inconsistently by the standards of Economic Man, but people will "irrationally" borrow money at a high rate of interest to avoid breaking into capital that is earning a lower rate (Harris & Laibson, 2001), lest the boundary between capital and spending money accounts be weakened.

To some extent personal rules can produce a pattern of preferences like Economic Man's, although the result will be more brittle than is usually imagined. Conventional Economic Man continually evaluates his prospects with an exponential discount curve. Intertemporal bargaining among hyperbolically influenced selves may arrive at a similar exponential curve as a personal rule to resolve their conflict (see Ainslie, 1991), but such a rule will not produce the same properties as a spontaneous preference that obeys an exponential function:

For one thing, evaluation by personal rules causes individual choices to be evaluated as precedents, often to a greater extent than as goods in their own right. Since value as a precedent does not depend on intrinsic value, but rather on the scope of the generalization future selves are apt to make from this case, a person's evaluation process will become somewhat legalistic. At some point this pattern becomes what clinicians call compulsive. It is arguably what existential philosophers have complained of as an "idealistic orientation," an "inauthentic" personal style (Ellenberger, 1958).

A second problem is that a single failure can lead to a sudden collapse of intertemporal cooperation, analogous to a stock market panic, which in addictionology is familiar as the "abstinence violation effect" (Curry, Marlatt, & Gordon, 1987). For some time after such a collapse, intertemporal cooperation in related areas may be possible only by abandoning the specific area where the failure occurred, leading to a circumscribed symptom, an impulse against which you are helpless. You *cannot resist* the urge to have a cigarette when nervous or to panic when speaking in public, and your belief that you cannot do so stabilizes the boundary between where your will is effective and where it is not.

A third problem is that the bargaining situation creates a motive to avoid information that might cause cooperation to collapse. If you have a strict rule against cruelty and one day you are cruel anyway, you will experience a sharp fall in whatever depends on your expectations of being a kind person unless you can keep from catching yourself. There are many ways to avoid catching yourself, described by psychodynamic therapists under terms like denial, suppression, rationalization, and that most unconscious block to information, repression. The whole Freudian underworld may arise not, as the psychoanalysts believed, from the painfulness of information *per se*, but from damage-control maneuvers in intertemporal bargaining (see Ainslie, 1982).

Finally, the increased effectiveness that well-marked criteria give to the bargaining process means that relatively concrete goals will have an advantage over subtle ones. Thus repeated choices of a countable reward such as money display valuations that are apparently exponential, as I described, while single trial experiments or real life choices among other kinds of goods elicit hyperbolic discounting. However, the person may find herself trapped by personal rules that seem necessary to avoid dangerous impulses, but that narrow her character pathologically. Miserliness and anorexia nervosa seem to be examples shaped by needing to be sure that you do not see yourself wasting money or obeying hunger (see Ainslie, 2001, pp. 143-160). "Buy only what you need" or "eat only what you will be content with in retrospect" are rules that are too nonspecific, too easy to evade by rationalization for someone very worried about her basic impulsiveness.

In summary, willpower as a solution to the intertemporal bargaining problem introduces four unintended effects that have obvious clinical implications: The more a person makes her choice-making systematic by inferring categories of future reward from her choices, the more brittle her behavior will become. She will seek goods not in proportion as they are rewarding, but insofar her choosing them satisfies criteria for intertemporal cooperation, a pattern that gets called compulsive. She may also develop areas of circumscribed failure, either time-limited, as with binges (anorexia not sustained becomes bulimia), or durable, as a character weakness (someone resigned to morbid obesity may still have a strong will in other areas). She may even rely upon the coexistence of regular lapses to escape from rules that are too narrow for her own longest range interest, a maneuver that may make addictions especially resistant to treatment. She will thus sometimes display preferences inscrutable to conventional utility theory, choosing options that she says she does not want, buying means to avoid choosing these goods but also undermining these means, and keeping herself from being able to report a substantial portion of what she is doing.

Vicarious Experience and Altruism: An Example of a Subtler Utility-Maximizing Problem

Hyperbolic discounting can now be said to be well established, but its implications have only begun to be explored. It predicts not only intertemporal bargaining and the cooperation that creates willpower, but a number of other motivational patterns that make no sense in terms of conventional utility maximization. I will illustrate these with the much-discussed question of what motivation there can be for vicarious experience-- in the extreme, for the altruism that leads a person to sacrifice what seem to be her own interests for another.

Although sociobiologists have argued that it is adaptive, up to a point, for a species to have its members sacrifice themselves for each other (Sober & Wilson, 1998), the behavioral sciences have been at pains to find a credible mechanism. It has even been argued that failure to find such a mechanism has made economists become less altruistic (Frank et.al., 1993). Arguments that altruism is a means of inducing reciprocity in its beneficiaries or is the product of a personal rule (Rachlin, in press) are defeated by counterexamples of generosity to passing strangers or by young children, and in any case

contradict the intuition that in many cases the raw feel of generosity is desirable.

The argument for how hyperbolic discounting makes altruism a primary good involves the interaction of several of its properties, and can only be sketched here (see Ainslie, 1995, and 2001, pp. 161-197). The foundation is the problem of premature satiation: There are many activities in which reaching peak intensity too fast reduces the total satisfaction that can be derived from them-- that is, where hasty consumption wastes appetite. Familiar examples include snacking to stay full, premature ejaculation, and reading ahead in a novel. Hyperbolic discounting implies that in activities where this is true people will be impelled toward the SS satisfaction at the expense of a LL reward pattern of gratifying the same appetite. We will get full satisfaction only where we can, by foresight or willpower, impose an obstacle that slows or defers consumption (Figure 4).

Figure 4A

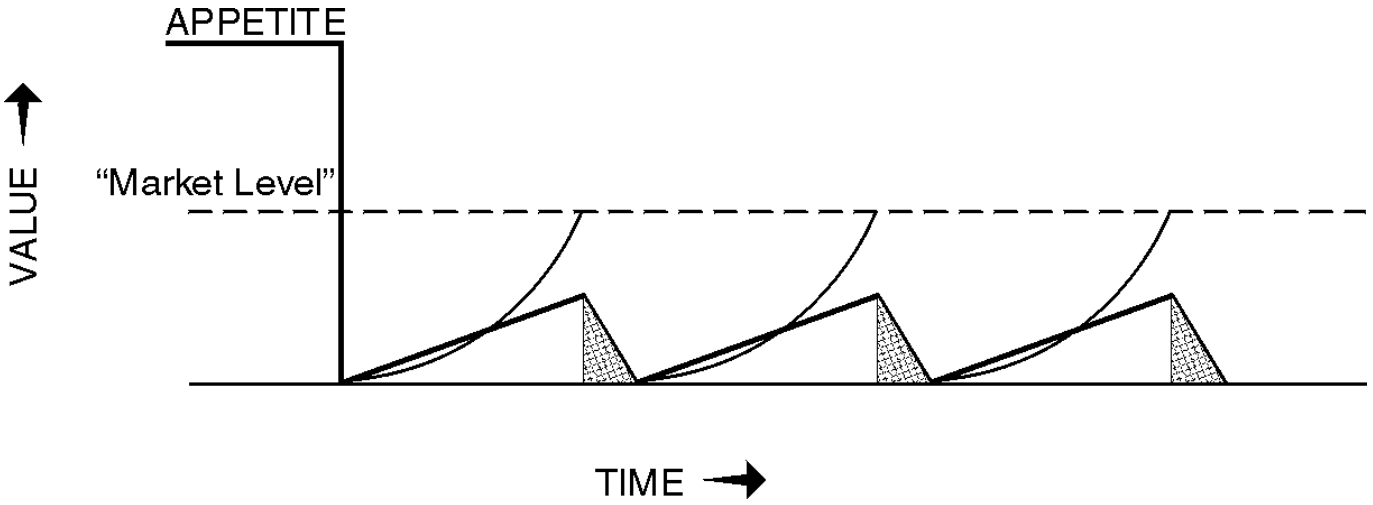

Figure 4A. Cycles of growing reward potential (depicted schematically as rising straight lines) and actual consumption (gray areas) leading to satiety. Consumption begins when discounted value of expected consumption reaches the competitive market level. Hyperbolic discount curves of the total value of each act of consumption decline with delay from its anticipated onset (right to left as delay increases).

Figure 4B

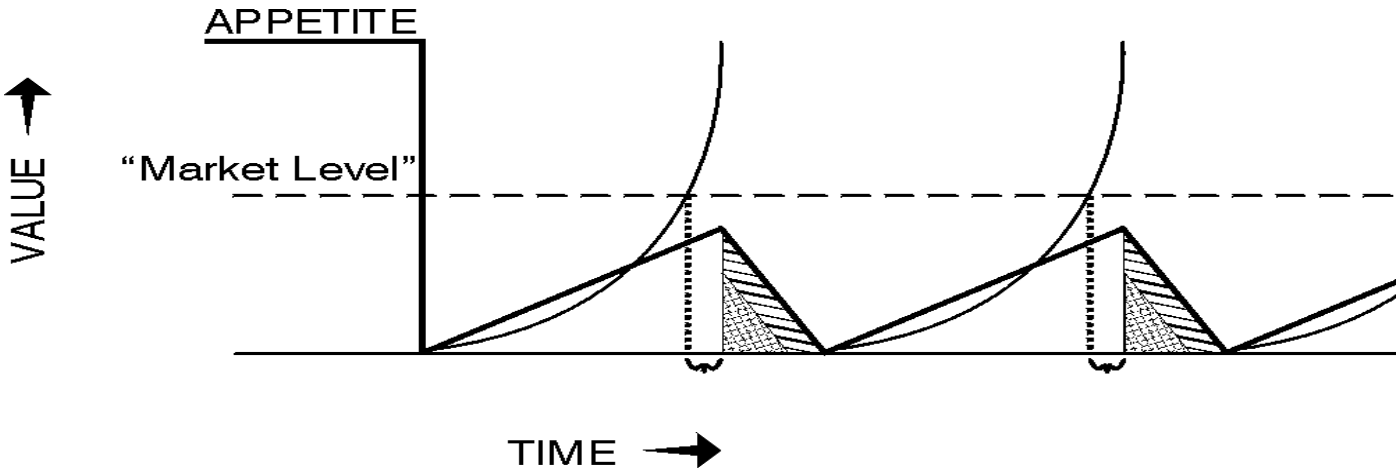

Figure 4B. Increased reward (stripes) resulting from increased appetite when there is an obligatory delay in the moment of starting consumption from the moment of choice ("{" brackets); the choice to consume occurs when the discounted value of the delayed consumption reaches the market level.

Emotion may be the most important area where premature consumption impairs reward.[2] Most human satisfaction is emotional, in the developed world at least, and emotions are available at any time without specific releasing stimuli. As any actor knows, they can be cultivated voluntarily, but doing so at will soon dissipates them prematurely; this is because the overvaluation of immediate reward causes attention to rush ahead to the climax of any familiar experience, producing a SS thrill instead of a LL one. Thus emotional rewards, although available without fixed stimuli, are actually constrained by a limitation that is outside of our arbitrary control, and this constraint exists precisely because of hyperbolic discount curves: Maximal satisfaction from emotional rewards depends on their deferral and the consequent buildup of appetite for them; hyperbolic discount curves create a relentless urge to harvest these rewards prematurely. Therefore, unless people peg their emotions to occasions that are both optimally unresponsive to their current wishes and optimally surprising, their emotional lives will have the highly satiated quality of daydreams.

The important question is how you can thus peg an emotion. At the most basic level it should happen by simple selection: All emotions not occasioned by cues that are outside of your voluntary control will satiate rapidly; as this limitation becomes familiar, you will not bother to begin them (cf. Frank, 1988, on the analogous social selection of sincere emotions). This learning process is apt to happen in nonhuman animals as well as people. Neurophysiology has recently suggested an important source of cues that are

[2] I have argued elsewhere that negative emotions like fear and mixed emotions like anger are not unrewarding, but lure us into participating in them by an attractive component that is temporally mixed with a deeper, unrewarding one-- that the motivational picture of these emotions is that of a rapidly cycling addiction (Ainslie, 2001, pp. 51-61). Thus they are more complicated than positive emotions, but can be treated like them for the purposes of the present argument.

externally generated but salient to our emotionality: We seem to be primed to register others' behaviors in the parts of our brain where we govern our own (*mirror neurons*-- Preston & deWaal, in press). But these vicarious experiences, being initiated from outside us, are much less subject to premature satiation. This process, too, is apt to happen widely in nonhuman animals, perhaps more in the social species than in others. A refinement of this process is the mental construction of actual models of another individual's experience, "if I were her," which occurs in even very young children and in some but not all great apes, and is increasingly studied as "folk theory of mind" or "*verstehen*" (Davies & Stone, 1995; Povinelli et.al., 1999). Such models will provide the most effective cues for pacing our own emotionality; that is, they are the most apt to provide occasions for entertaining emotions that are surprising, on the one hand, but integrated with our own familiar experience on the other. Thus there is probably a hierarchy of sophistication for occasioning emotions. At whatever level, the hyperbolic discounting hypothesis is that emotion is limited not by the availability of stimuli to elicit it but by the premature satiation of emotions that are not paced by salient and surprising external events. We invest in other people's experiences because they provide both salience and surprise. But the emotions we entertain "vicariously" when putting ourselves in others' shoes are actually occurring in our brains, and must reward us in the same way as our "own."

This theory of vicarious experience as a primary good leaps a long way ahead of agreed-upon facts. Its elements-- that emotion is a reward-dependent behavior, that emotion is limited by premature satiation, and that vicarious experience gets its value by providing surprise-- are predicted by hyperbolic discounting but by no means established. I have detailed the theory here to provide a motivationally consistent alternative to the more circuitous ways that theorists are now explaining altruism. It is at least *possible* to understand the vicarious experience that motivates altruism as the primary good that intuition says it is, without abandoning a strict utility-maximizing model. I also wanted to illustrate how far-reaching the implications of hyperbolic discounting may be, using altruism as one of several equally important examples (see Ainslie, 2001, pp. 143-197).

Future Directions

Corrections based on hyperbolic discounting might allow rational choice theory and corresponding economic theory to account in unified fashion for the greater part of the anomalies that have confronted it. Beyond just explaining existing anomalies, hyperbolic discounting generates hypotheses about fundamental motivational processes, the exploration of which reached dead ends in behavioral science a generation ago. Questions like why pain is not the symmetrical opposite of pleasure, whether involuntary processes such as emotions and appetites are motivated like other behaviors, and whether a second principle of behavioral selection such as classical conditioning is necessary to account for organisms' participation in their own aversive experiences, are re-opened at their roots. Their answer will require not only controlled experimentation, which has limited power with recursive phenomena, but also methods for analyzing what people know from experience beyond mere introspection. The example that I have found most

promising is the kind of thought experiment that is well developed in the philosophy of mind.

Conventional utility theory still has little hold on the areas of need that are the most moving to people in a society that is highly satiated with material goods, as exemplified by the need for surprise and for the vicarious experience of other people. I believe that this hold cannot be extended by the application of utility theory as it is. Rather we should view the anomalies that have been discerned thus far as penetrations of an underlying hyperbolic principle into classical utility theory. The classical theory is a largely coherent but narrowly bounded body of behavior shaped by a particular human situation: the competition of farsighted people for dominance in a marketplace of tangible goods. That model is to motivational science as a whole what classical physics is to relativistic physics, a special case within a more universal but less accessible system. The job that beckons is to deduce and then test the properties of the larger system.

References

Ainslie, George (1974) Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior* 21, 485-489.

Ainslie, George (l982a) A behavioral economic approach to the defense mechanisms: Freuds energy theory revisited. *Social Science Information* 21, 735-779.

Ainslie, George (1991) Derivation of "rational" economic behavior from hyperbolic discount curves. *American Economic Review* 81, 334-340.

Ainslie, George (1992) *Picoeconomics: The Strategic Interactionof Successive Motivational States within the Person*. Cambridge: Cambridge U.

Ainslie, George (1995) A utility-maximizing mechanism for vicarious reward: Comments on Julian Simons "Interpersonal allocation continuous with intertemporal allocation" *Rationality and Society* 7, 393-403.

Ainslie, George (2001) *Breakdown of Will*. New York, Cambridge U.

Ainslie, Goerge and Monterosso, John (in press) Building blocks of self-control: Increased tolerance for delay with bundled rewards. *Journal of the Experimental Analysis of Behavior.*

Bickel, Warren K., Odum, Amy L., and Madden, Gregory J. (1999) Impulsivity and cigarette smoking: Delay discounting in current, never, and ex-smokers. *Psychopharmacology* 146, 447-454.

Boring, E.G. (1950) *A History of Experimental Psychology* New York: Appleton-Century-Crofts.

Bratman, Michael E. (1999) *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge, UK, Cambridge U.

Carrillo, Juan D. (unpublished) Self-control, moderate consumption, and craving. Universite Libre de Bruxelles, March 1999.

Carver, Charles S. and Scheier, Michael F. (1990) Principles of self-regulation: Action and emotion. in *Handbook of Motivation and Cognition: Foundations of Social Behavior, v. 2* (E. Tory Higgins and Richard M. Sorrentino, eds.) New York, Guilford, pp. 3-52.

Curry, S., Marlatt, A., and Gordon, J.R. (1987) Abstinence violation effect: Validation of an attributional constructwith smoking cessation. *Journal of Consulting and Clinical Psychology* 55, 145-149.

Darwin, Charles (1872/1979) *The Expressions of Emotions in Man and Animals.* London: Julan Friedman Publishers.

Davies, Martin and Stone, Tony (1995) *Folk Psychology: the Theory of Mind Debate.* Oxford, UK: Blackwell.

Ellenberger, Henri F. (1983) A clinical introduction to psychi-atric phenomenology and existential analysis. in R. May, E.Angel, and H. Ellenberger (eds.), *Existence: A New Division in Psychiatry and Psychology.* N.Y.: Basic Books,Inc., pp. 92-124.

Frank, Robert H. (1988) Passions Within Reason, New York: W.W. Norton and Company.

Frank, Robert .H., Gilovich, Thomas and Regan, Dennis (1993) Does studying economics inhibit cooperation? *Journal of Economic Perspectives* 7, 159-171.

Green, Leonard, Fry, Astrid, and Myerson, Joel (1994) Discounting of delayed rewards: A life-span comparison. *Psychological Science* 5, 33-36.

Harris, Christopher and Laibson, David (2001) Dynamic choices of hyperbolic consumers. *Econometrica* 69, 535-597.

Kavka, Gregory (1983) The toxin puzzle *Analysis* 43, 33-36.

Kirby, Kris N. (1997) Bidding on the future: Evidence against normative discounting of delayed rewards. *Journal of Experimental Psychology: General* 126, 54-70.

Kirby, Kris N., and Guastello, Barbarose (2001) Making choices in anticipation of similar future choices can increase self-control. *Journal of Experimental Psychology: Applied* 7, 154-164.

Klein, B. and Leffler, K.B. (1981) The role of market forces in assuring contractual performance. *Journal of Political Economy* 89, 615-640.

Korobkin, Russell and Ulen, Thomas S. (2000) Law and Behavioral Science: Removing the Rationality Assumption from Law and Economics, *California Law* Review 88, 1051-1144.

Laibson, David (1997) Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics* 62, 443-479.

McClennen, Edward F. (1990) *Rationality and Dynamic Choice.* .New York: Cambridge.

Mazur, J.E. (1987) An adjusting procedure for studying delayed reinforcement. in M.L. Commons, J.E. Mazur, J.A. Nevin, andH. Rachlin, (eds.*), Quantitative Analyses of*

Behavior V: TheEffect of Delay and of Intervening Events on ReinforcementValue, Hillsdale, N.J.: Erlbaum.

Mazur, James E. (2001) Hyperbolic value addition and general models of animal choice. *Psychological Review* 108, 96-112.

Metcalfe, Jane and Mischel, Walter (1999) A hot/cool-system analysis of delay of gratification: Dynamics of willpower. *Psychological Review* 106, 3-19.

Monterosso, John and Ainslie, George (1999) Beyond Discounting: Possible experimental models of impulse control. *Psychopharmacology* 146, 339-347.

Monterosso, John, Ainslie, George, Toppi Mullen, Pamela, and Gault, Barbara (2002) The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma. *Journal of Economic Psychology* 23:4, 437-448.

Povinelli, Daniel J., Bierschwale, Donna T., and Cech, Claude G. (1999) Comprehension of seeing as a referential act in young children, but not juvenile chimpanzees. British Journal of Developmental Psychology 17, 37-60.

Preston, Stephanie B. and de Waal, Frans B. M. (in press) Empathy: Its ultimate and proximate basis. *Behavioral and Brain Sciences.*

Rachlin, Howard (in press) Altruism and selfishness. *Behavioral and Brain Sciences.* Russell, J.Michael (1978) Saying, feeling, and self-deception. *Behaviorism* 6, 27-43.

Schelling, Thomas C. (1960) *The Strategy of Conflict*. .Cambridge, Mass: Harvard University Press.

Shefrin, H.M. and Thaler, R.H. (1988) The behavioral life-cycle hypothesis. *Economic Inquiry* 26, 609-643.

Simon, Julian L. (1995) Interpersonal allocation continuous with intertemporal allocation: Binding commitments, pledges, and bequests. *Rationality and Society* 7, 367-430.

Sober, Elliott and Wilson, David Sloan (1998) *Unto Others: the Evolution and Psychology of Unselfish Behavior*. Harvard U.

Sorensen, Roy A. (1992) *Thought Experiments*. New York,Oxford.

Stearns, Peter N. (1994) *American Cool: Constructing aTwentieth-Century Emotional Style*. New York: New York U.

Vuchinich, Rudy E. and Simpson, Cathy A. (1998) Hyperbolic temporal discounting in social drinkers and problem drinkers. *Experimental and Clinical Psychopharmacology* 6, 292-305.