Recursive Self-Prediction in Self-Control and its Failure

George Ainslie

Veterans Affairs Medical Center, Coatesville

Published in *Preference Change: Approaches from Philosophy, Economics, and Psychology*
Till Gruene-Yanoff and Sven Ove Hansson, eds.
Springer, 2009, pp. 139-158

This material is the result of work supported with resources and the use of facilities at the Department of Veterans Affairs Medical Center, Coatesville, PA, USA. The opinions expressed are not those of the Department of Veterans Affairs or of the US Government.

Abstract

The combination of human foresight and the discounting of delayed events in a hyperbolic curve is all that is needed to explain the learning of higher mental processes *from the bottom up*. These processes are selected by delayed rewards.insofar as they counteract the over-valuation of imminent rewards that is also predicted by hyperbolic discounting. For instance, these processes come to interpret repeated, similar choices as moves in an intertemporal bargaining game resembling an iterated prisoner's dilemma. Perception of current choices as test cases for cooperation in such a game recruits the extra motivation experienced as willpower. Lines seen as criteria for such tests may be experienced as beliefs rather than resolutions. The chance that shifts of self-prediction may cause radical swings of motivation makes choice unpredictable from just knowing the person's prior incentives, even by the person herself; the resulting introspective uncertainty is arguably the subjective basis of freedom of will. A similar kind of recursive self-prediction explains how surges of emotion or appetite can be occasioned by symbols that convey no information about the availability of external rewards.

Text

There is a basic tendency for humans and nonhuman animals to change their preferences from larger, later (LL) rewards to smaller, sooner (SS) rewards in the absence of new information about their availability or proximity. This tendency is best called *impulsiveness*, although the term has also been used trivially to describe spontaneity or poor motor inhibition. I will first review work presented elsewhere on the hyperbolic shape of the function that describes devaluation of delayed reward: the problem that maintaining consistent choice poses for evolution, and how this shape is apt to govern both impulsive changes of preference and methods of limiting these changes. I will then expand on my previous suggestion that the most important of these methods, the interpretation of current choice as a predictor of future choices, exemplifies a phenomenon that can be inferred not only in conscious impulse control, but in such basic experiences as freedom of will, emotion, appetite, belief, and character.

The observation of *recursive self-prediction*—self-prediction that is fed back to the ongoing choice process-- is limited by its inaccessibility to controlled experiment, but this phenomenon is predictable from experiments that are not only controlled but precisely quantitative; and it can be tested by other, less direct means. In my view its existence challenges the conventional assumption that preferences govern only voluntary choices, and that preferences are in turn governed by an overarching faculty of will. It opens the possibility that a broader array of mental processes than is usually imagined competes in a common marketplace of reward, and that self-control and other higher mental functions can grow from the bottom up through interaction in this marketplace. Recursive self-prediction probably mediates a great deal of human experience.

Hyperbolic discounting poses a problem in adaptiveness

Impulsiveness is fully explained only by the finding that reward-seeking organisms devalue prospective events in a hyperbolic function (Ainslie 1975, 2001), which describes value as a simple inverse proportion of delay:

$$\text{Value} = \frac{\text{Value at no delay}}{[\text{Constant} + (\text{Impatience factor} \times \text{Delay})]}$$

This function predicts temporary preference for SS over LL rewards when the SS rewards are closer. Such hyperbolic discounting has stood up in repeated testing (Green & Myerson, 2004; Kirby, 1997), which has only suggested modifications to the basic formula in the possible addition of an exponent to the denominator of the function (Green& Myerson, 2004), a feature that does not affect its implications for motivating preference change. Alternative explanations involving the classical conditioning of appetites (Lowenstein, 1996; Laibson, 2001) or shifting cognitive frames (Rubinstein, 2003; Trope & Liberman, 2003) actually require hyperbolic discounting as an underlying mechanism (Ainslie, in press).

Hyperbolic discounting raises the obvious question of how people ever avoid switching their preferences toward SS rewards as they come close—that is, achieve the consistent behavior that is the norm of rational choice theory (RCT; Herrnstein 1990; Boudon 1996) and the requisite for success in financial markets. This is not an issue for nonhuman animals, in which long range planning has been shaped by natural selection in the form of specific hardwired instincts. Animals mate, defend territory and hoard food for the winter not to ensure offspring, maximize resources, and prevent future starvation, but to gratify current urges. Even chimpanzees can wait only a few minutes to get increased amounts of favorite foods (Beran and Evans 2006). However, the necessity of coding long range rewards into lifelong instincts greatly limits a species' ability to learn new environmental contingencies. When an instinctive method of hoarding is cracked by interlopers, countermeasures will not appear for many generations if a new instinct has to evolve. It would clearly be more efficient for an organism to try different hoarding strategies on the basis of the long-term results they produce, so that failure would cause the loss of only the effort of a particular strategy, not a whole organism. There do exist examples where nature has given nonhumans an ability to learn from long-delayed consequences. In bait shyness, for instance, an animal can learn to avoid a taste that has been followed by sickness hours later, but the range of possible learning is narrow: The cue has to be a taste rather than a visual appearance, and the consequence has to be nausea rather than somatic pain (Garcia *et.al.* 1974). You might think that a mechanism of more flexible choice among outcomes of varied delays would have evolved much earlier; but the hyperbolic discount curves that move animals to promptly obey instincts make long range intertemporal choice potentially disastrous; SS rewards will tend to dominate LL ones. Given any ability to short-circuit the instinctual mating process, for

instance, animals become vigorously autoerotic, as anyone who has visited the monkey house in a zoo can testify.

Hyperbolic discount curves have created a major pitfall for the evolution of flexible intelligence, to the extent that there is a serious question of how these curves evolved. There are two possible rationales, one of them unlikely. It could be argued that behaviors such as mating and fighting benefit the species at the expense of the individual's long range interest, so groups that discounted urges for them hyperbolically were selected; however, individuals' awareness of their long range interests evolved long after the form of the discount curve did. The more likely, and simpler, answer is that hyperbolic curves are a previously harmless manifestation of a universal psychophysical principle: that changes in a sensory quantity are perceived as a proportion of the baseline quantity-- the Weber-Fechner law as applied to delay or some correlate of delay (Gibbon 1977). Such proportionality is also described by a hyperbola. Hyperbolic curves were harmless until organisms became intelligent enough to manipulate their sources of reward. As long as reward is controlled by the contingencies with which a species' instincts evolved, prompt obedience to those instincts will be the individual's best bet. Conversely, the hyperbolic shape may be what has limited the evolution of intelligence, but is so basic to the structure of motivation that it cannot be replaced at this late stage. Imaginative humans have learned to divorce pleasure from its original adaptive purposes to an enormous extent, mating, eating, and behaving in general to get pleasure rather than to increase reproductive fitness. Great skill at taming nature does not correlate (positively, at least) with the production of children in modern society. In combination with hyperbolic discounting, skill makes the individual dangerous even to herself. Control over reward lets her take her life in her hands, with enormous motivation to waste her resources—addiction is a human phenomenon. And when competing for these resources with an individual who has learned to evaluate them consistently over time-- a human skill that I will discuss presently-- she is at risk of becoming a money pump—someone who sells her winter coat every spring and buys it back at a higher price every fall (Cubitt and Sugden 2001).

The combination of intelligence and hyperbolic discounting clearly poses a risk, but one that some people seem to overcome fairly well. How does someone with hyperbolic discount curves sometimes manage to keep to the plans that her own foresight dictates? Furthermore, this question is not the greatest one posed by the hyperbolic discount function. Although motivational inconsistency is the first issue that comes to mind in contemplating hyperbolic curves, fundamental assumptions about the self come into question soon after.

Hyperbolic discounting creates motivation for developing higher mental functions

The conventional idea of the self is that of a unitary executive that is entirely able to command some subordinate faculties—motor behavior, for instance, both current and future—and totally unable to control many other important processes such as appetites, emotions, and involuntary behaviors, especially the "negative" processes that would not be chosen deliberately. This self is substantial, impenetrable, and exempt from the strict

laws of physical causality: It is felt to be substantial in the sense that it comprises more than the set of its motives, and has a form of inertia-- the tendency of a choice to remain in place from the mere fact of having been made. It seems impenetrable in not being susceptible of analysis into simpler components. And although it can cause actions through its function of will, incompatibilist doctrines of free will state that it is not bound by causes acting upon it in turn (Clarke 2003). However, the hyperbolic shape of the basic discounting curve raises the question of whether any of this is necessarily so. Motivational theory can break free of the early behaviorists' model, the Skinner-box-writ-large that was so unlike the experience of complex choice (*e.g.* Skinner, 1948), and contemplate higher mental functions with very different properties: held together only by motivation, analyzable with game theory, and predictive of the experience of free will while remaining strictly within the chain of causality as conventionally understood. If mental processes are shaped by a single, or common, selective factor that decays hyperbolically from the time of choice to the time of reward, it turns out to be fairly easy to model a self with these features.

Start with the concept of *value*, defined as the property of inducing behavioral selection: The functional effect of an event's value is the tendency of an organism to select a mental process that is followed by the valued event. A valued event is *a reward* (whereas the selective influence itself is just *reward*, without the definite article-- potentially confusing, but it follows existing usage). The simplest model of choice is that an organism generates an array of options and selects the one that has the greatest expectable reward, discounted for delay and uncertainty. The precise way that options are generated and compared does not matter here, but it might be imagined to be something like Edward Tolman's concept of vicarious trial and error (1939), the rehearsal of each contemplated course of action before actually adopting one. Such a process has lately been observed physiologically in the rat hippocampus— The neurons subtending possible paths become active alternately until one path wins and choice moves forward (Johnson and Redish 2007). We would expect options that never win to eventually drop out of the array, so that reward affects not only the selection of a process but also the endurance of this process as an option.

If prospective reward were discounted in a function that produced consistent choice—*exponentially*—experience would affect subsequent choices only by changing the individual's expectations of delay and uncertainty. In a farsighted organism a faculty of self would be needed only to estimate what string of options chosen consistently, would produce the greatest aggregate of discounted, expected reward over time. Selves would be mere calculators, and the process of choice would be determined by the estimated contingencies of reward, "throughput" as J. M. Russell called it (1978). Naturally theorists who imagine such a process of choice see the need to find extrinsic motives for impulsiveness such as sudden appetites driven by association, and for selves that perform impulse control by transcending mere motivation. However, given that prospective events are evaluated *hyperbolically*, options—or, more precisely, the mental processes that try to obtain these options-- must compete with each other on the basis not only of each option's delay and uncertainty but also of their relative delays. Put another way, values shift relative to one another as a function of elapsing time, and thereby introduce

an additional element of uncertainty to each option, even if the option is certain to be obtained if chosen. Mental processes that pursue contradictory options may each survive in an individual's array of choices because none dominates the others at all times. With a hyperbolic discount function, maximizing prospective discounted reward at one moment no longer "makes" a choice. To keep getting the reward that originally shaped it, the mental process pursuing that reward has to add means of staying chosen. The mind then functions as a population, not because it contains contradictory options—these would exist as well if rewards were discounted exponentially—but because the processes rewarded by these options have incentives to predict and forestall each other. This is the implication of hyperbolic discounting that lets it predict more than impulsiveness; it shapes the basic relationships that can ramify to form a self from the bottom up.

To reach fruition an option must promise not only the greatest discounted prospective reward of a current array of options *if it were certain*; it must also promise to withstand challenges by competing options that may look better before it comes to fruition. Its value is adjusted for the uncertainty that this very competition introduces. This problem can be demonstrated in, and sometimes solved by, a pigeon: If a peck on a red key leads to an SS reward, and no peck to an LL reward, and if an earlier peck on a green key simply keeps the red key from subsequently appearing, some birds learn to peck the green key (Ainslie 1974). This is impulse control of the simplest sort, and does not require the subject to have any functional knowledge of why pecking the green key leads to greater prospective discounted reward as of that moment. The pigeons that learn this kind of precommitment could be said to have foresight of a sort for the time periods involved—a matter of seconds—but not self-awareness. Even the most foresighted problem-solvers—people—have had limited success in devising impulse-control devices. External devices such as guardians and restricted bank accounts have limited availability and scope; diverting attention works only in the short run; and cultivating or inhibiting influential emotions (the psychoanalysts' reaction formation or isolation of affect) has significant costs. The external device that people have used most has been the influence of other people, sometimes in the form of physical controls—parents' control of their children, governments' enforcement of laws—but more robustly in other people's ability to give or withhold occasions for emotion (see Ainslie 1995). However, these social commitments also have limitations, especially as we devise increasingly cosmopolitan societies. They become dangerous when you meet a person who wants to exploit you, a likelihood that increases with the number of people you meet; they give way when a whole group has the same impulse, a phenomenon that Jan Huizinga described as prevalent in the late middle ages (1924) but which still recurs in the form of "war fevers" and the "madness of crowds" generally; and they are useless against impulsive behaviors that can be concealed. The device that has best combined strength and flexibility has been another one altogether, which the individual exercises autonomously; it has been nebulous from the viewpoint of motivational science.

Recursive self-prediction provides a mechanism for will

An ability to stabilize one's own choice for one's own welfare was gradually differentiated from conscience in the sixteenth century, became a fashion in the

seventeenth, and has been the subject of many theories since, often under the name of *will*. The early psychologists began cataloguing its properties (Sully 1884, pp. 630-670; James 1890, pp. 486-592), but the lack of externally observable markers led it to be stigmatized as an unscientific concept, and discussion of it dried up almost completely as the twentieth century unfolded (sketched in Ainslie 2001, p. 202, note 12). The will was held to be as inscrutable as the self (*e.g.* Pap 1961), from which it has not been clearly bounded. The absence of analytic discussion of a process that is so central to human functioning has been striking, suggesting a hesitation, even a queasiness, about putting mortal fingers on it, the kind of discomfort that some religions have had about naming their deity. However, from a scientific standpoint the main obstacle to analyzing the will has been the lack of a motivational rationale for it.

"Will" has been used to name the process by which intention is connected to motor movement, and the sense of ownership that someone has of her actions (Wegner 2002), but its most important meaning is the process that restrains impulses (See Ainslie 2004). The philosophers and psychologists who have given advice about the will over the centuries have discerned several attributes, most notably a basis in choosing according to principle rather than according to the particulars of the current circumstance. The power of this abstract idea to reduce actual impulsiveness is puzzling from the viewpoint of RCT, which depicts people as naturally consistent to begin with; but it is predicted by the hyperbolic discount function, given only two conditions: that the cumulative discounted value of a series of expected rewards is roughly additive, and that a person's expectation of getting the whole series can be made contingent on her current choice without physical commitment. The additivity condition has been verified experimentally (Mazur 2001; Kirby 2006), as has its implication that subjects will show greater preference for LL over SS rewards when choosing a whole series at once instead of singly. This increase in patience has been found in students choosing between amounts of money, and of pizza; subjects who chose every week for five weeks between a smaller, immediate amount and a larger amount a week later were much more likely to choose the SS amount than subjects who had to make their choice for all five weeks at once, on the first week (Kirby and Guastello 2001). The same pattern has been observed in rats choosing amounts of sugar water (Ainslie and Monterosso 2003). The replication of this finding in animals shows that the increase in patience comes from the properties of the basic, presumably hardwired discount function itself, rather than depending on cultural suggestion or on an effect of total amount on patience (an effect seen only in humans-- Green *et.al.* 2004).

The second condition— that a person's mere perception of her current choice as a test case predicting how she will choose in the future can bundle series of choices together— does not lend itself to experimental test. However, the dependence of large expectations on current test cases is a common intuition. The cost to a dieter of eating a piece of chocolate is clearly not a detectable gain in weight, but her loss of the expectation that she will stick to her diet. Uncontrolled observations of several kinds support this intuition: The lore on willpower mentions a role for a bad precedent in reducing willpower (*e.g.* Bain 1859/1886, p. 440); when Kirby and Guasello suggested to their student subjects that each weekly choice predicted how they would make subsequent choices, they moderately increased the subjects' preference for LL alternatives (2001);

and vulnerability to perceived lapses can be modeled by interpersonal bargaining games (Monterosso *et.al.* 2002). However, the best way to test the original intuition is to sharpen it by a device popular in the philosophy of mind, the thought experiment. I have argued that a small number of selected thought experiments yield a valid rejection of the null hypothesis-- that contingent self-prediction is unnecessary for volition (Ainslie 2001, pp., and in press). Direct observation will be impractical for the foreseeable future; even functional magnetic imaging (fMRI), which has localized the components of many motivational processes (Cardinal 2006), cannot show the semantic content of such processes.

With our present observational abilities we can only follow out the implications of hyperbolic discounting, and test what we see against the familiar properties of volition: An individual with foresight who notices the predictiveness of present choices should develop processes that look very much like a will and a self by experience alone, without their being supplied *ex machina* by a homunculus: A self-aware hyperbolic discounter will learn to take into account the existence of other relevant processes that have been shaped differently by different temporal relations with the same reward center(s). Processes that are congenial to each other will cohere into the same process. Contradictory ones will treat each other as strategic enemies. Ineffective ones will cease to compete at all. Thus hyperbolically discounted reward will create what is in effect a population of reward-seeking processes that group themselves loosely into *interests* on the basis of common goals, just as economic interests arise in market economies. The choice-making self will have many of the properties of an economic marketplace, with a scarce resource—access to the individual's limited channel of behavior—bid for with a common currency—the prospect of reward. The logic of repetitive bargaining games will create regularities within this marketplace, including reliable support for those farsighted processes that can predict and act early to forestall or foster processes will be strongly motivated by imminently available rewards. Maintenance and change of choice will be governed by *intertemporal bargaining*, the activity in which reward-seeking processes that share some goals (e.g. long term sobriety) but not others (when to have drinks) maximize their individual expected rewards, discounted hyperbolically to the current moment. This *limited warfare* relationship is familiar in interpersonal situations (Schelling 1960, pp. 21-80), where it often gives rise to "self-enforcing contracts" (Klein and Leffler 1981) such as nations' avoidance of using a nuclear weapon lest nuclear warfare become general. In inter*personal* bargaining, stability is achieved in the absence of an overarching government by the parties' recognition of repeated prisoner's dilemma incentives. In inter*temporal* bargaining *personal rules* arise through a similar recognition among the successive motivational states of an individual, with the difference that a future state is not motivated to retaliate, as it were, against past states that have defected. In the intertemporal case the risk of future states' loss of confidence in the success of the personal rule, and their consequent defection in their own short term interests, will present the same threat as the risk of actual retaliation. These contingencies can create a will without an organ, serving a self without a seat, just as the "will" of nations not to use nuclear weapons seems to be guided by an invisible hand.

In this way will can grow from the bottom up, through the selection of increasingly sophisticated processes by elementary motivations. In many depictions from Descartes onward the will has the appearance of a canoeist steering through rapids—using skill and foresight to ride forces much stronger than itself, but still something made of different stuff, a spirit, a homunculus. The intertemporal bargaining process grows the canoeist from the stuff of the rapids, different in skill and foresight but subject to the same motivational forces, and in fact developed by those forces. It is when the canoeist learns to include her own future tendencies as part of the currents she must anticipate that a pattern recognizable as a self develops. As with many natural patterns, this mechanism is most recognizable where pathology exaggerates it, for instance in obsessive-compulsive personality disorder and encapsulated areas of dyscontrol (Ainslie 2001, pp. 143-160). Here I will focus just on the way that recursive self-prediction permits the leap from current to canoeist, that is, from strict causality to the experience of free will.

When the incentives for alternative choices are closely balanced, small changes in the prospects for future cooperation swing the decision between cooperation and defection. In that case an assumption about the direction of the present choice will be a major factor in estimating future outcomes. But this estimate in turn affects the probability that the present choice will be in that direction. Thus the decision process is recursive-- not tautological, but continuously fed back like the output of a transistor to its own input. If the person's predictions about her propensity to make the choice in question are at all open, this feedback process may play a bigger role in her decision than any given incentive, external or internal. For instance, a dieter faces a tempting food, guesses that she will be able to resist it, applies the consequences of this guess to the expected reward contingencies as an increase in the likelihood that she will reap the benefits of her diet, and thus has more to stake against the temptation. Then she discovers a credible loophole and thereby incurs a fall in her expectation of a successful diet because of the chance she will try the loophole and not get away with it—that is, the chance that she will subsequently judge her choice to have been a lapse, thus reducing the stake against further lapses. This fall may be so great as to make the expected values of lapsing vs. trying to diet about equal, until some other consideration tips her self-prediction one way or the other. Such a process is not subtle conceptually, but it eludes any calculation based only on the contingencies of reward, and buffers the person's decision against coercion by these contingencies. Thus it can be argued to generate the experience of exercising free will (Ainslie 2001, pp. 129-134). Furthermore, such an explanation allows us to characterize free choices better than saying that they are too close to predict. After all, many behaviors are quite predictable in practice and are still experienced as free. What becomes crucial is the person's belief that a given choice depends on this self-prediction process, however she has come to represent this process to herself.

Diets and resolutions are examples of consciously constructed personal rules, with clearly defined conditions as to what kinds of choice are members of the relevant bundle, and criteria for which choices are cooperations and which are defections. However, once an individual has discovered that her current choice gives her predictive information about her future choices, even choices that are not governed by resolutions are apt to be influenced by this information to a greater or lesser extent. This influence will be largely

nameless, or be hidden in seemingly disparate processes with names like force of habit, being true to yourself, or even responding to beliefs about the world. True, this recursive influence may sometimes serve purposes other than deterring impulses. For instance, I may habitually gather tasks to take to the office near my front door the day before I leave, either (1) so I can find them easily when I'm in a hurry, or (2) so as to keep myself from putting off doing them,. Purpose (1) makes this activity a coordination game without a conflict of interest between myself currently and in the future; purpose (2) recognizes a repeated prisoner's dilemma, designed to coerce my future self by making any act of procrastination set a precedent. The difference may be perceptible in whether or not I experience the habit as having force: A coordination game can be changed without compunction if, say, a more convenient mnemonic device comes along. Change in a repeated prisoner's dilemma for what looks like momentary convenience may produce an unaccountable feeling of unease, which is a sign that I have suspected the choice was really a lapse of intertemporal cooperation.

Recursive self-prediction accounts for sudden appetites and emotions

There is no reason why recursive self-prediction should be limited to conscious volition. There are many common experiences where a mental process that is under marginal control is influenced by signs of how it is progressing. 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 generally 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 common in behaviors that you recognize to be only marginally under voluntary control: summoning the courage to perform in public or face the enemy in battle, recall an elusive memory, sustain a penile erection, or, for men with enlarged prostates, void their bladders. William James went as far as to say that we feel an emotion only when we detect somatic manifestations of it—a theory that has been shown to be overstated (Rolls 2005, pp. 26-30), but which may well describe how quasi-voluntary processes are accelerated or modulated.

But how can processes that are more or less involuntary fit the same recursive pattern as will? The hyperbolic shape of the discount curve supplies an answer, by allowing us to broaden our concept of reward, and hence of motivation. The existence of an internal marketplace for positive incentives has long been assumed by utility theorists, economists foremost among them. Recently neurophysiologists have reiterated the necessity of

recognizing such a marketplace (Shizgal and Conover 1996); that is, a mechanism by which all substitutable processes can be weighed against each other In a marketplace model many diverse processes compete for a limited channel of attention on the basis of a common dimension of selectability, such that an relative increase in this dimension for an act of game-playing, say, or charity, can lead it to be selected over an act of food consumption, while a relative decrease for the game or charity could lead the consumption to be selected. However, only desirable processes are usually imagined to compete directly with one another. Intuition has dictated that aversive processes participate only negatively in this marketplace—that they are introduced by a non-market process and have their effect only by making subsequent escapes rewarding. We use the words "reward" or "utility" for a property that is deliberately sought, and different words such as "urgency" or "vividness" for a property that seems to demand attention without being desirable, yet the latter terms also imply positive motivation—motivation that impels you into an experience. The notion that aversive processes are directly selectable along the same dimension as desirable ones seems to depart from intuition, but part of the problem is linguistic. If we stop equating rewardingness with desirability—the property that lets something be deliberately sought—and define it more basically as the property that makes whatever process it follows tend to be repeated, we can avoid having to explain the force of aversive experiences with a second, non-market process.

Examples such as nausea, rage, and fear are processes that are usually thought of as unmotivated—what is the incentive to be nauseated?—but rather imposed on the individual by a reward-independent process such as classical conditioning. An opposing view has long pointed out that the selective factors in classical conditioning--unconditioned stimuli—invariably have incentive value as well as the power to condition, and has suggested that conditioning is a form of reward-governed learning (Hilgard and Marquis 1940; Donahoe *et.al.* 1993), The difficulty with this theory is that the incentive value of unconditioned stimuli is often negative, that is, that they select for processes which the individual is motivated to avoid. The frequent vividness of the negative emotions has seemed to demand a second kind of selective factor, which rewards attention while deterring physical approach. In the conventional model, pain, fear, grief, anger, and presumably nausea are imposed in reflex fashion either by innately programmed turnkeys or by stimuli that have been associated with such turnkeys. However, conditioned attention and reward-seeking participation look very much alike. The reward-responsiveness of negative emotions can sometimes be discerned in the cases where they have come under voluntary control: Sometimes people have learned to pay attention to a painful stimulus without emitting the emotion-like response that makes pain aversive ("protopathic" as opposed to "epicritic" pain—Sternbach 1968), or to withhold a fear response to stimuli that have been provoking it (Clum 1989). Anger may feel imposed by a circumstance, but everyone has sometimes experienced the competition between "bothering" with an anger and carrying on the activity that it threatens to spoil—a competition that is apt to turn on the rewardingness of the alternative activity. Indeed, anger shares many psychometric and neurophysiological properties with the more obviously positive emotions, such as increased optimism, heuristic as opposed to reflective cognitive processing, and left as opposed to right frontal cortical activation (Lerner and Tiedens 2006).

I have argued elsewhere that the hyperbolic discounting of reward permits the modeling of negative, positive, and mixed emotion-like processes by the cyclic mixture of reward and subsequent inhibition of reward (Ainslie 1992, pp. 100-114; 2005). To summarize briefly: Just as a cycle of binge and hangover attracts and then repels behavior over a period of days, and as nail-biting or tics attract choice only when they are possible within seconds (cf. Berridge's "wanted but not liked" behaviors, 2003; also Peciña *et.al.* 2006), so an urge to panic or attend to a traumatic memory may be "satisfied" only for a split second before its aversive effect is felt. Such an urge attracts attention but deters physical approach, exactly the effect of conditioned *negative* emotions. For motivated *positive* emotions, the question is why they would not lead to autistic self-reward. The brief answer is that hyperbolically-based preference for SS over LL emotional experiences should motivate premature satiation unless this activity is limited to adequately rare occasions; I shall say more about this presently. Even daydreams must include obstacles if they are to escape complete habituation. Finally, the mixture in *mixed* emotions is not a weighing of two opposite valences—which would lead to neutrality—but rather the perception that a strongly motivated emotion will bring just enough aversiveness to make its desirability from a distance ambiguous.

The ability of negative incentives to compete in the internal marketplace on the basis of a single selective factor—reward—permits a wide range of involuntary processes to be brought into this marketplace. The set of reward-seeking behaviors will comprise all internal processes to the extent that they compete with one another for expression. In particular, emotion becomes a form of behavior. The sensation of being cut or burned offers an opportunity for the emotion of protopathic pain—an opportunity that is hard, but not necessarily impossible, to refuse. The sensation of tossing in a boat offers the opportunity for nausea, the perception of loss offers the opportunity for grief or anger, and so on. Many processes remain outside of this set, for instance the competition of a muscular extension reflex with an opposing contraction reflex; and many processes take part in the set only partially. Cardiac contractions and peristalsis are somewhat autonomous, in that they will occur regardless of that an individual is thinking or feeling, but regrets, daydreams, plans for dinner, awareness of an itch, and excruciating pain all compete with each other, however unequally. The more one occurs, the less room the others have to occur. Even cardiac contractions and peristalsis can be brought into this marketplace to a limited extent, when sensations from them come to attention or when activity in a market member (*e.g.* fear) raises or lowers their activity; but their core functioning remains outside the market. Sometimes a pathologic phenomenon shows that a seemingly autonomous activity must have been occupying a small space in the market, as when loss of the urge to breathe—a motivation not usually noticed-- impairs respiration ("Ondine's curse;" Kuhn *et.al.* 1999). Sometimes deliberate learning enlarges the market-responsive component of autonomous activities, as when cardiac contractions or peristalsis come under the control of hatha yoga or biofeedback (Basmajian *et.al*. 1989). The boundaries of the internal marketplace are not sharp and may be variable to some extent, but they clearly include much more than the set of voluntary activities or the set of desirable activities. The point for the present discussion is that not only deliberate

but also involuntary reward-seeking processes should be affected by recursive self-prediction.

The value of the marketplace model can be seen in the example of sudden craving. Conditioned appetite has been proposed as the explanation of the sudden cravings that people develop for food or drugs when they encounter reminders of them, particularly when the people are trying to avoid consuming them (Loewenstein 1999; Laibson 2001). However, in laboratory examples of conditioning, conditioned stimuli lead to responses only when they predict imminent consumption. If a conditioned stimulus (CS) occurs or begins well before its unconditioned stimulus (UCS) is due, subjects learn to estimate the delay and emit the conditioned response (CR) just before the UCS (Kehoe *et.al.* 1989; Savastano *et.al.* 1998; see Ainslie, in press). The alternative that hyperbolic discounting makes possible is that appetites are reward-dependent processes, and that their sudden arousal in the absence of any increased availability of their objects is an attempt to make consumption of these objects more likely. The logic is as follows: Reward-dependent processes compete for acceptance on the basis of the current discounted value of the prospective reward for these processes. An appetite arises when an individual perceives the opportunity for consumption that can be made either more rewarding or more likely by this appetite; appetite may serve not only to prepare for consumption, but to make consumption more likely. In examples of elicited appetite in the laboratory, the timing of consumption is necessarily controlled by the experimenter. In daily life, by contrast, goods that might be consumed impulsively are available much of the time, and their consumption is limited by a person's decisions. If a random appetite increases the rewardingness of a prospective object, it increases the likelihood that the person will consume the object, which will induce further appetite in preparation for the possible consumption. This is a positive feedback system, driven by the person's recursive self-perception of the likelihood that appetite will be enough to make her decide to consume the object. It has the same math as Russell's seasickness, the expectation of vomiting that confirms itself.

A sudden spike of appetite could thus come from the existence of positive feedback conditions. These conditions may obtain whenever the person's consumption is determined mainly by her choice about a readily available consumption good, but are apt to have the strongest effect when there is weak-to-moderate resolve not to consume: Where a person is not trying to restrain consumption she will keep appetite relatively satisfied; where she is confident of not consuming regardless of appetite she will not expect appetite to lead to consumption. In neither of these cases will appetite be rewarded by motivating consumption. In a recovering addict or restrained eater, by contrast, cues predicting that she might lapse could significantly increase the likelihood of lapsing. There will still be constraints on the motivation for an appetite—in modalities where unsatisfied appetite brings hunger pangs or withdrawal symptoms these will be deterrents; and appetite without a limited occasion will extinguish (see Ainslie 2001, pp. 166-171)—but the explosive appetite that so often ends people's efforts at controlled consumption can be understood as a motivated process that has sought to do exactly that.

This model depends on the hyperbolic shape of the discount curve, since an individual with consistent preferences over time would have no short range motive to undermine her own resolutions, or indeed any long range motive to make resolutions in the first place. Given such motives and some self-awareness, recursive self-prediction can be expected to punctuate consistent behavior with fits and starts of appetite.

Beliefs may arise through recursive self-prediction

In a model of the individual as a population of reward-dependent processes, facts can be seen as what constrains the search for reward. The experience of being constrained by facts is called belief. In highly imaginative organisms such as humans relatively little reward comes from current sensory experience, or even from the prospect of any sensory experience that is so imminent that it demands attention. Most of our significant prospects are relatively distant, complex, and subject to interpretation. These prospects reward us as occasions for current emotion, in competition with other occasions such as the vicarious experience of another person, or pure fantasy ("make-believe"), as well as sensory experience itself. As I mentioned above, an occasion paces reward most effectively when it is relatively infrequent, which in practice means that it must be governed by contingencies other than the immediate rewarding potential of the emotion, and connected to the emotion in some way that lets it stand out from other possible occasions.[1] To keep from paling into a daydream the joy of winning must be occasioned by new information, specific to a person or project or sports team or even fictional story to whom or to which you have already given importance. Similarly an occasion for panic must have some connection to pain or loss, but will be less apt than joy to pale, because of your avoidance of such occasions.

Although there are many possible rationales for making occasions unique—a longstanding practice or a myth shared by an entire culture or even good fiction-writing technique-- the simplest way of being unique is to be factual. The scenarios that are instrumental in changing the real world are apt to also be those that compete best in the marketplace at the current moment, but not necessarily because of the prospect of experiencing their practical results; they have hedonic impact beyond this prospect as occasions for emotion that are more unique than make-believe (see Lea and Webley 2006). However, the motivational impact of make-believe can be amplified to a comparable level by reducing the freedom to choose alternatives; commitment to the outcomes of particular fictional scenarios in online fantasy projects such as Second Life may yield emotions as imperative as "realistic" activities such as day trading. What makes Second Life more powerful than a video game is the extent to which it is a single consensual project that cannot be cheaply abandoned for another one. Fictional works may achieve this uniqueness by becoming cultural icons—as Schelling describes for the death of Lassie (1986)—or even by an individual's single-minded devotion to one immutable set of outcomes.[2] Such examples elevate "make believe" to made beliefs—commitments to occasions for emotion that are divorced from instrumental effectiveness in the real world but which are binding enough to have the same hedonic impact. If belief is basically the experience of being constrained by facts, the irreplaceable

ingredient is not the descriptive truth of the facts but rather the emotional cost of escaping them.

The role of the perceived facts themselves is often unclear. We have a strong tendency to discern facts underlying constraints, but to the extent that practical instrumentality is not important, the facts that we identify may serve more as labels for particular constraints than as predictors of external rewards. Perhaps the most important source of these constraints that do not come from physical limitations is intertemporal bargaining. One example is the way that people experience the non-predictive cues that lead to appetites, described above. As with all processes for which reward is freely available a cue is needed only to give occasion, that is, to select one moment from among many to make a focused bid for expression. Often the environment is a strong selective factor—coming upon food or a loss or a confrontation—but often the occasion comes from a mere reminder or symbol. Even then, a cue that leads to a feeling one time becomes more likely to do it the next time, because it increasingly stands out from other available occasions as the association is repeated. Soon it will be experienced as "the reason for" the appetite or emotion. That is, even when the first occasion was a random stimulus its evocativeness will come to seem like a fact of the external world.

Personal rules supply another important example of perceived factuality that comes from intertemporal bargaining. The very volatility of recursive self-prediction means that people will be apt to cling to rationales for truces, that is, to lines between do-able self-control and futile efforts. Again uniqueness is valuable-- here the quality of being a *bright line*, a boundary between conflicting interests that cannot be shifted without inviting more shifts. A recovering alcoholic has an available bright line between some drinking and no drinking at all. A dieter has only lines laid down by diets, which are much dimmer in the sense that they are more replaceable by other authors' lines that do not stand out any less. Lines like these, which are the criteria of personal rules, are often experienced as facts, the more so the brighter they are. For instance, recovering alcoholics have long believed that they have a biological susceptibility that causes a single drink to lead to irresistible craving; but it has been shown experimentally that it is the belief that they have had a drink of alcohol, not the alcohol itself, that is followed by craving (Maisto *et.al.* 1977).

Our inherited instinct for disgust turns upon mostly ambiguous stimuli in the modern world. The process of recursive self-prediction creates the belief that some things *are* dirty, occasions for disgust, and others *are* clean. Accepted authorities may alter boundaries between them, as when the mania for cleanliness early in the twentieth century followed the discovery of germs, and is said to be resurgent now after the discovery of new diseases (Ashenburg 2007, pp. 239-289); but often our instinct for disgust seems to be controlled by rituals of just sufficient difficulty, adjusted for how strong our individual instinct is to begin with. For instance, there is no scientific reason to avoid touching urine, your own or someone else's. Only a single, tropical disease is transmitted through human urine, and that not through simple contact.[3] Nevertheless urine is universally assumed to *be* a contaminant, a belief that waxes and wanes, however, inversely with the difficulty of avoiding it. Parents of young children

experience a sudden reduction in their belief, and people on camping trips are not generally bothered by the impossibility of washing after urination. The British colonial army in the nineteenth century could carry only limited equipment on bivouac, and used the same trough for washing in the morning that they had used as a urinal the previous evening (Farwell 1985). Reduction in the behavior of urine-avoidance drives a reduction of the belief that it is needed, a change that is even more apparent in the converse situation of germ-phobics: Avoidance of a new kind of contact, with a doorknob, say, sets a precedent of treating doorknobs as contaminated, and is in danger of making the person grasp them only through a handkerchief in the future. The most effective treatment of this and other phobias is behavioral—graded exercises in which the patient acts as if the fear were not true (Marks 1997).[4] Of course the same person may do lip service to very different beliefs, but the actual constraint she is under is the behavioral boundary established by recursive self-prediction.

The belief that you have found a bargain can be instantly rewarding. The hunt for bargains produces the pleasure in many kinds of shopping, whether "compulsive" or not. However, maintenance of this belief requires behavior that is consistent with it. If you have stocked up on food at a good price or bought a concert series at a discount, you may face an incentive to eat the food when you are tired of it or attend a concert you do not expect to enjoy in order to avoid recognizing a loss. And yet you may be fully conscious of the unpleasant prospect. The belief in the bargain is really a personal rule for playing a game, the wins in which occasion emotional reward that is related only tangentially to the reward of tasting the food or listening to the concert. The relationship is that the prospect of this consumption authenticates the bargain-hunting as an instrumental activity rather than a mere game, even though, once so authenticated, the bargain-hunting is a self-sufficient source of reward and has requirements that sometimes contradict those of optimally consuming the ostensible reward.

Another personal rule that masquerades as a belief is a performer's self-confidence. A performer can be defined broadly as anyone whose activity can be ruined by a loss of nerve—comedian, acrobat, public speaker, even warrior or lover. The belief has the form, "I *am* able (funny, nimble, persuasive…)," but it depends on the behavior of not fleeing, literally or emotionally, from the activity. Such flight, incisively named "flopsweat" by comedians, has the same incentives as any other kind of panic—the insubstantial relief of gratifying an urge that nevertheless beckons insistently. A large component of the self-confidence is the expectation that you can avoid panic, which adds a stake to the avoidance but perversely, in this case, increases the urge to panic for that very reason; thus self-confidence is particularly prone to the positive feedback phenomenon. Performers often find that they need additional resolutions: avoiding defensiveness, not playing for applause, not copying past work, and other formulae for resisting short range rewards; these again may take the form of beliefs: "The audience doesn't matter" or "I'm doing this for art's sake."

The difference between a conscious resolution and a constraint that is experienced as a fact may sometimes lie in how much of your prospective reward is at stake in the relevant choices. Conversely, you may increase the prospect at stake and thus your motivation for

self-control by interpreting your personal rule as a response dictated by a belief. A person who resolves to be vegetarian to conserve the earth's resources does not face a strong incentive never to backslide; a person who believes that animals *are* fellow souls and eating them *is* murder will be committed much more strongly, to the point even that she will begin to experience disgust rather than pleasure at the thought of eating meat (see related studies by Paul Rozin, *e.g.* Rozin *et.al.* 1997). A single lapse will have much broader implications than it would for the environmentalist, perhaps instilling doubt about her basic character.

An increased stake in a personal rule will increase the ease of following it, but also increase the loss if you do not. The increase in stake could come either from a long history of success, or the perception of this rule as a key component of a broader and more important rule—against cruelty, dishonesty, or perversion, for instance. At some point you will cease to perceive the rule as a resolution and experience it instead as a trait of your character: "I *am not the kind of person* who…" can kill, is sneaky or mean spirited, or might have a disgusting paraphilia. This is a stake that is threatened by even a single lapse, greatly increasing your motive to avoid catching yourself lapsing. It is arguably the maneuver discovered by John Calvin, which gave the early Protestant burghers their legendary ability to defer consumption (Weber 1904/1958); see also my discussion in Ainslie 2001, pp. 134-139): If any sin is a sign that you are among those predestined to damnation, it makes a sin much more important than just a single failure of good works. If you have such a belief, a lapse faces you with a choice among 1. modifying your belief, but thereby giving up its committing power; 2. accepting the prospect of damnation, which in motivational terms is probably the same as #1; or 3. rationalizing so as not to classify the behavior as a lapse—usually the least costly solution, and probably the greatest source hypocrisy where deceiving others is not a factor. Although it is conventional to distinguish character traits from behaviors that are merely habitual, and is thus natural to distinguish the "self-signaling" that will not tolerate lapses from less consequential self-prediction (Prelec and Bodner 2003), they are just different zones on a continuum.

Conclusions

Recursive self-prediction is the expectable consequence of hyperbolic discounting in self-aware individuals. It is inaccessible to controlled experimentation, but offers a parsimonious model of several otherwise puzzling human phenomena:

- Higher mental functions, exemplified by will, do not require unified faculties but rather can be seen as intertemporal bargaining skills that become included in reward-seeking mental processes to the extent that they lead these processes to be better rewarded.
- "Free will" describes the experience of predicting your choices in a way that also modifies these choices, making them unpredictable from a knowledge of the original incentives but not excepting them from literal causality.
- Involuntary processes such as appetite and emotion may be selected by the same mechanism that selects deliberate choices, the recursive prediction of which

explains sudden eruptions following "conditioned" stimuli, without our having to attribute special properties to the association process.

- Belief can be seen as the recognition of constraints on choice, which include incentives that are recruited through self-prediction but that are experienced as facts. The perception of commitment to some kinds of self-control as a character trait increases the extent of this commitment.

The inadequacy of previous bottom-up theories in explaining higher mental processes may have been due to their depiction of motivation as a linear product of the person's incentives. Recursive self-prediction is not an exceptional process, but is probably present in most human intentionality. To paraphrase physicist Stanislaw Ulam, "the study of non-linear motivation is like the study of non-elephant zoology."

References

Ainslie, G. 1974. Impulse control in pigeons. *Journal of the Experimental Analysis of Behavior* 21: 485-489.

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

Ainslie, G. 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Cambridge U.

Ainslie, G. 1995. A utility-maximizing mechanism for vicarious reward: Comments on Julian Simon's "Interpersonal allocation continuous with intertemporal allocation" *Rationality and Society* 7: 393-403.

Ainslie, G. 2001. *Breakdown of Will*. Cambridge U.

Ainslie, G. 2004. The self is virtual, the will is not illusory. *Behavioral and Brain Sciences* 27: 659-660.

Ainslie, G. 2005. Précis of *Breakdown of Will*. *Behavioral and Brain Sciences* 285., 635-673.

Ainslie, G. in press. Hyperbolic Discounting versus Conditioning and Framing as the Core Process in Addictions and Other Impulses. In *What Is Addiction?*, ed. D. Ross, H. Kincaid, D. Spurrett, and P. Collins. MIT.

Ainslie, G. and Monterosso, J. 2003. Building blocks of self-control: Increased tolerance for delay with bundled rewards. *Journal of the Experimental Analysis of Behavior* 79: 83-94.

Bain, A. 1859/1886. *The Emotions and the Will*. Appleton.

Basmajian, J.V. et.al. 1989. *Biofeedback: Principles and Practice for Clinicians.* 3d Ed. Baltimore: Williams &Wilkins.

Bennett, A. 1918. *Self and Self-Management*. George H. Doran.

Beran, M. J. and Evans, T. A. 2006. Maintenance of delay of gratification by four chimpanzees (Pan troglodytes.; The effects of delayed reward visibility, experimenter presence, and extended delay intervals. *Behavioural Processes* 73: 315-324.

Berridge, K.C. 2003. Pleasures of the brain. *Brain and Cognition* 52: 106-128.

Boudon, R. 1996. The "rational choice model:" A particular case of the "cognitive model." *Rationality and Society* 8: 123-150.

Cardinal, R.N. 2006. Neural systems implicated in delayed and probabilistic reinforcement. *Neural Networks* 1277-1301.

Clarke, Randolph. 2003. *Libertarian Accounts of Free Will*. Oxford.

Clum, G.A. 1989. Psychological interventions vs. drugs in the treatment of panic. *Behavior Therapy* 20: 429-457.

Cox, F. E. G. 1993. *Modern Parasitology: A textbook of Parasitology. 2d Edition.* Wiley.

Cubitt, R.P. and Sugden, R. 2001. On money pumps. *Games and Economic Behavior* 37: 121-160.

Darwin, C. 1872/1979. *The Expressions of Emotions in Man and Animals*. Julan Friedman.

Donahoe, J.W., Burgos, J.E., and Palmer, D.C. 1993. A selectionist approach to reinforcement. *Journal of the Experimental Analysis of Behavior* 60: 17-40.

Farwell, B. 1985. *Queen Victoria's Little Wars.* Norton.

Garcia, J., Hankins, W. and Rusiniak, K. 1974. Behavioral regulation in the milieu interne in man and rat. *Science* 185: 824-831.

Green, L. and Myerson, J. 2004. A discounting framework for choice with delayed and probabilistic rewards. *Psychological Bulletin* 130: 769-792.

Green, L., Myerson, J., Holt, D.D., Slevin, J.R., and Estle, S.J. 2004. Discounting of delayed food rewards in pigeons and rats: Is there a magnitude effect? *Journal of the Experimental Analysis of Behavior* 81: 39-50.

Herrnstein, R. J. 1990. Rational choice theory: necessary but not sufficient. *American Psychologist* 45: 356-367.

Hilgard, E.R. and Marquis, D.G. 1940. *Conditioning and Learning*. Appleton-Century.

Huizinga, J. 1924. *The Waning of the Middle Ages*. St. Martin.

James, W. 1890. *Principles of Psychology.* New York: Holt.

Johnson, A. and Redish, A.D. 2007. Neural ensembles in CA3 transiently encode paths forward of the animal at a decision point. *Journal of Neuroscience* 12: 483-488.

Kehoe, E. J., Graham-Clark, P., and Schreurs, B.G. 1989. Temporal patterns of the rabbit's nictitating membrane response to compound and component stimuli under mixed CS-US intervals. *Behavioral Neuroscience* 103: 283-295.

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

Kirby, K.N. 2006. The present values of delayed rewards are approximately additive. *Behavioural Processes* 72: 273-282.

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

Kirby, K. N. 2006. The present values of delayed rewards are approximately additive. *Behavioural Processes* 72: 273-282.

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

Kuhn, M., Lutolf, M., and Reinhart, W.H. 1999. Ondine's Curse. *Respiration International Review of Thoracic Disease.* 663: 265.

Laibson, D. 2001. A cue-theory of consumption. *Quarterly Journal of Economics* 66: 81-120.

Lea, S.E.G. and Webley, P. 2006. Money as tool, money as drug: The biological psychology of a strong incentive. *Behavioral and Brain Sciences* 29: 161-209.

Lerner, J.S., and Tiedens, L.Z. 2006. Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making* 19: 115-137.

Loewenstein, G.F. 1996. Out of control: Viscera influences on behavior. *Organizational Behavior and Human Decision Processes* 35: 272-292.

Loewenstein, G.F. 1999. A visceral account of addiction. In *Getting Hooked: Rationality and Addiction,* ed. J. Elster and O.-J. Skog. Cambridge U.

Maisto, S., Lauerman, R. and Adesso, V. 1977. A comparison of two experimental studies of the role of cognitive factors in alcoholics drinking. *Journal of Studies on Alcohol* 38: 145-49.

Marks, I. 1997. Behavior therapy for obsessive-compulsive disorder: A decade of progress. *Canadian Journal of Psychiatry* 42: 1021-1027.

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

Monterosso, J., Ainslie, G., Toppi- Mullen, P., and Gault, B. 2002. The fragility of cooperation: A false feedback study of a sequential iterated prisoner's dilemma. *Journal of Economic Psychology* 234.: 437-448.

Pap, A. 1961. Determinism, freedom, moral responsibility, and causal talk. In *Determinism and Freedom in the Age of Modern Science* , ed. Sidney Hook. Collier.

Pecina, S., Smith, K.S., and Berridge, K.C. 2006. Hedonic hot spots in the brain. *The Neuroscientist* 12: 500-511.

Prelec, D. and Bodner, R. 2003. Self-signaling and self-control. In *Time and Decision: Economic and Psychological Perspectives on Intertemporal Choice, ed.* George Loewenstein, Daniel Read, and Roy Baumeister., pp. 277-298. Russell Sage.

Rolls, E.T. 2005. *Emotion Explained.* Oxford U.

Rozin, P., Markwith, M., and Stoess, C. 1997. Moralization and becoming a vegetarian: The transformation of preferences into values and the recruitment of disgust. *Psychological Science* 67-73.

Rubinstein, A. 2003. "Economics and psychology"? The case of hyperbolic discounting. *International Economic Review* 44: 1207-1216.

Russell, J.M. 1978. Saying, feeling, and self-deception. *Behaviorism* 6: 27-43.

Savastano, H. I., Hua, U., Barnet, R. c. and Miller, R. R. 1998. Temporal coding in Pavlovian conditioning: Hall-Pearce negative transfer. *Quarterly Journal of Experimental Psychology* 51: 139-153.

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

Schelling, T.C. 1986. The mind as a consuming organ. In *The Multiple Self, ed.* J. Elster. Cambridge U.

Shizgal, P., and Conover, K. 1996. On the neural computation of utility. *Current Directions in Psychological Science* 5: 37-43.

Skinner, B.F. 1948. Superstition in the pigeon. *Journal of Experimental Psychology* 38: 168-172.

Sternbach, R.A. 1968. *Pain: A Psychophysiological Analysis*. Academic.

Sully, J. 1884. *Outlines of Psychology.* Appleton.

Tolman, E. C. 1939. Prediction of vicarious trial and error by means of the schematic sowbug. *Psychological Review* 46: 318-336.

Trope, Y., and Liberman, N. 2003. Temporal construal. *Psychological Review* 110: 403-421.

Weber, M. 1904/1958. *The Protestant Ethic and the Spirit of Capitalism.* Charles Scribners Sons.

Wegner, Daniel M. 2002. *The Illusion of Conscious Will*. MIT.

---

[1] For aversive emotions these requirements are less stringent, since a person's motivated avoidance of them keeps them uncommon; also, for evolutionary reasons aversive emotions seem to habituate less than pleasurable emotions.

[2] A fictional but credible example is the hero of Robert Coover's *The Universal Baseball Association* (1968) who has invested his emotions so much in a single, long-continuing fantasy baseball game that the randomly determined outcomes have the impact of facts (Ainslie 1992, p. 313-315).

[3] One strain of schistosomiasis, a parasitic infection, is spread by infected urine in bathing sites (Cox 1993).

[4] Compare Arnold Bennett's advice for curing "fussiness" by deliberately acting contrary to fussy beliefs about yourself as soon as you identify them (1918, p. 80).