rency among economists, the idea is not completely consensual. The phenomenon of preference reversals is certainly an empirically robust fact. But as Rubinstein (2003) and other economists have recently remarked, the adoption of hyperbolic discounting does far more than just modeling the psychological phenomenon that the present has a special status. "It assumes the maximization of a utility function with a specific structure and as such it misses the core of the psychological decision-making process. Thus, I find it to be no more that a minor modification to the standard discounting approach" (Rubinstein 2003, p. 1215).

The objection raised by Rubinstein is that the same sort of evidence provided by preference reversals can also reject hyperbolic discounting as well; although most of these phenomena can be explained in terms of a decision procedure based on similarity relations, Rubinstein's own proposal constitutes an even more severe departure from RCT (e.g., transitivity is not satisfied). It has nevertheless the merit of delivering an account of other important choice anomalies that inspired the main theories of nonexpected utility.

Of course, a sophisticated theory of rationality cannot be based on an unmodified version of RCT. This is so in virtue of *normative* reasons alone. For example, some phenomena, like Ellsberg's, seem to require a normative relaxation of RCT of the sort advocated by Amartya Sen (2002) or Isaac Levi (1986). An adequate descriptive theory of choice should also be able to explain problematic patterns of choice of this kind. But it seems doubtful that the adoption of a particular functional form for utility can accomplish that. The so-called theories of nonexpected utility, which also relied on the use of special utility curves, have been relatively successful in explaining only a limited range of recalcitrant phenomena (e.g., Ellsberg's puzzle cannot be explained by appealing to the special utility and probability curves of prospect theory; especial and questionable ad hoc hypotheses have been used instead by Fox and Tversky [1991] – see Arló-Costa and Helzner [2005] for a critical appraisal of this account).

In Chapter 4 of Ainslie's book, hyperbolic discounting is applied via ingenious arguments to develop accounts of emotions, pain, and other aversive behavior. The extension requires the neat separation of the notions of reward and pleasure. Therefore, the theory is quite different from the standard hedonic articulations of rational choice. In addition, the model presents an account of emotions as "pulled" by reward, which seems, prima facie, controversial.

But the bulk of the book is devoted to showing that one can develop a model of the will as emerging from a process of intertemporal bargaining among units called *interests*. This is one of the most imaginative and interesting parts of the book. Perhaps the most charitable account of interests is as time slices of the self.

Stroz concluded in his seminal paper (Stroz 1955) that "the individual always decides what to do on the assumption that he has no authority over his future self" (p. 180). Since then, many philosophers have questioned this assumption by postulating that agents can bind themselves through the operation of their wills. This strategy *postulates* rather than explains intentionality. Ainslie suggests instead that when someone seems to be choosing according to principle (or by using a personal rule), what literally happens is that his successive selves form a repeated prisoner's dilemma relationship, which is solved in the manner of interpersonal bargainers. So the idea here is to explain the will away, rather than invoke it to explain commitment.

Ainslie concedes that the existence of the resulting internal feedback process is probably impossible to study via controlled experiments. Nevertheless, he claims that postulating hyperbolic discounting (and therefore recursive decision-making) solves some well-known philosophical conundra on intentionality and choice. I only have space here to focus on one of them: the so-called Newcomb's problem.

As originally formulated, Newcomb is a single-play situation and the puzzle is to determine what is *normatively* required in this case. Ainslie seems to think (citing Nozick 1969) that in this case RCT requires choosing both boxes (p. 134). But Nozick's argument insists that if one maximizes "evidential" expected utility (EU), one should choose one box. Dominance is invoked to rationalize the two-box solution. Later on, deviant versions of RCT (causal versions) have been developed in order to articulate the latter solution (Joyce 1999). Presystematically, Ainslie seems to be a two-boxer (p. 137). But this conclusion is deeply controversial (see, e.g., Levi 1975; Meek & Glymour 1994), and it does not seem to be based on using hyperbolic discounting or recursive decision-making.

In addition, the author argues that the temptation to hedge on a personal rule could be modeled as a problem with the same arithmetic structure as a repeated version of Newcomb. In his model, he argues, diagnostic acts are also causal acts and one can then explain the sequential analogue of the "evidential" solution in the single-play case (cooperating). But this does not seem to explain why Ainslie defends the solution analogous to mutual defect in the one-shot case of Newcomb. And this is the gist of Newcomb's problem.

Ainslie's very ambitious idea is to open the impenetrable black box of intentionality, which is then modeled as a sort of brokerage process. But this process is hard to dissect on account of its recursive nature. I do not think that the appeal to philosophical puzzles like Newcomb's adds credibility to it. Many of these puzzles share with Ellsberg's example the feature of being thought experiments designed to uncover *normative* inadequacies of RCT. And the theory presented in the book, like many other descriptive theories of choice, seems unable to deal with this type of problems. Ultimately, the plausibility of Ainslie's theory seems to rest on how well it fares in comparison to other attempts to "open the black box of decision."

The final chapters of the book (on the "dyscontrol" symptoms that can be induced by bargaining strategies) are full of fascinating insights. Probably, Ainslie offers in this book one of the most complete and theoretically unified theories of the will available today. The resulting overall picture is certainly quite impressive.

Three other motivational factors

Kent Bach

Department of Philosophy, San Francisco State University, San Francisco, CA 94132. **kbach@sfsu.edu** **http://online.sfsu.edu/~kbach**

Abstract: Ainslie uses his hyperbolic discount model to explain a dazzling array of puzzling motivational phenomena. In so doing, he assumes that the motivational force of a given option at a given time is directly proportional to its discount-adjusted reward as assessed at that time. He overlooks three other factors which, independently of the perceived reward, can affect motivational force.

Ainslie (2001) assumes that the motivational force of a given option at a given time is directly proportional to its discount-adjusted reward as assessed at that time. Evidently, he rejects any independent role for cognition in mediating or arbitrating competing rewards or in deliberating and deciding what to do when. Rather, he conceives of these interests as a population of quasi-independent agents engaged in tacit bargaining, each aimed at its own temporally discounted reward. He argues that the curves representing this discounting are "highly bowed," hyperbolic rather than exponential in form, thereby allowing for temporary reversals of preferences. He briefly mentions four other discounting patterns (p. 208), but, mathematically speaking, there are countless others consistent with temporal reversal. Indeed, perhaps the discounting takes different shapes for different rewards, and maybe the discounting is sometimes non-monotonic, as with highly unstable desires. But let's assume that Ainslie's contention that they are hyperbolic is not hype and that he is not taking us for pigeons.

https://doi.org/10.1017/S0140525X0523011X Published online by Cambridge University Press

The key idea is that hyperbolic discounting, by devaluing future rewards and punishments (negative rewards) proportionately to their delay, lets "utility theory move beyond its stalemate with cognitivism" (p. 38). So-called "dynamic inconsistency" is really a side effect of the fact that the discount curves for different rewards can cross: "people will naturally go for smaller, earlier over later, larger rewards. . . . Akrasia is just maximizing expected reward, discounted in highly bowed curves" (p. 39). In explaining how preferences can temporarily reverse, this simple model eliminates the apparent mystery of how we can act against our better judgment and against our "true" interests. Ainslie contends that hyperbolic discounting can explain a host of phenomena, including impulsiveness, addiction, compulsion, ambivalence, procrastination, and back-sliding. It can explain "the irony of smart people doing stupid things or having to outsmart themselves in order not to" (p. 27), by adopting "personal rules," cultivating good work habits, and making commitments that increase the cost of yielding to temptation. Consideration of future rewards doesn't take us outside the realm of reward and require higher-level judgments. There is just the ongoing competition among the rewards themselves.

It may seem an exaggeration to treat a person's different values as autonomous agents engaged in intertemporal bargaining with one another. Ainslie himself recognizes that his seemingly schizophrenic model of "the self as a population" (p. 39) makes it puzzling how, in a dog-eat-dog world of competing bargaining agents, "a marketplace of hyperbolically discounted choices [could] ever come to look like a single individual" (p. 40) rather than a kennel. What most worries me, though, is something else: At a given time, the motivational force of a desire (drive, urge, goal, value, or whatever you want to call the members of this population) is not a function merely of the reward associated with it, even as adjusted for the odds of success and the cost of failure and as temporally discounted.

Here are three other factors that can contribute to the motivational force of a particular desire: (1) the frequency and persistence of the desire's coming to mind, (2) the desire's degree of insatiability, and (3) its resistibility to the second-order desire to get rid of it. Each of these factors can vary even as the perceived reward of what is desired stays the same. For example, (1) something you want but deem of little importance can be more zealously pursued simply because the thought of it keeps occurring to you and capturing your attention. Playing another video game when you are trying to finish writing a commentary does not seem all that important, but its urgency is enhanced just because the thought of doing it keeps occurring to you. This doesn't make it seem like a better thing to do. Rather, you think you had better do it. What is rewarding is not playing the game but eliminating the clamoring desire to play it. Even worse, sometimes (2) the thought of playing the game does not go away after you play it. You did what you wanted to do, but now you want to do it again, just as if you hadn't done it in the first place. It's not that you want it more but that you want it again. In such a predicament, you may desire to make this desire go away, but (3) try as you may, it keeps rearing its head, keeping you from concentrating on that commentary. Now you're back to square (1), and vulnerable to (2) and (3) all over again.

You could have the opposite problem, say with a long-term project that requires irregular but frequent attention over time. As much as you value the ultimate reward of cultivating your garden, for example, (1) the thought of doing even a little puttering does not occur as frequently as it should. Not only that, (2) when it does occur and you act accordingly, the mere satisfaction of doing a little something makes you feel as though you have made significant progress even though you haven't. And, to make matters worse, (3) the thought of cultivating your garden resists staying in mind even when you want it to.

It might seem that these three factors are reducible to the perceived size of the desired reward. However, to suppose that would confuse the assessed size of the desired reward with these three independent dimensions of strength of the desire itself.

I have three further worries about the explanatory depth and breadth of the hyperbolic discount model. First, it ignores the distinction between something's being desired because it is rewarding and something's being rewarding because it is desired. Second, this model does not explain the magnitudes assigned to particular rewards in the first place. And third, the fact, if it is a fact, that discounting is hyperbolic itself cries out for an explanation.

Hyperbolas and hyperbole: The free will problem remains

Bruce Bridgeman

Department of Psychology, University of California–Santa Cruz, Social Sciences 2, Santa Cruz, CA 95064. **bruceb@ucsc.edu**
http://psych.ucsc.edu/faculty/bruceb/

Abstract: Hyperbolic theories have the fatal flaw that because of their vertical asymptote they predict irresistible choice of immediate rewards, regardless of future contingencies. They work only for simple situations. Theories incorporating intermediate unconscious choices are more flexible, but are neither exponential nor hyperbolic in their predictions. They don't solve the free will paradox, which may be just a consistent illusion.

Will a hyperbolic theory of reward discounting solve the persistent problems of the role of will in governing behavior? Ainslie (2001) makes a case for that idea, but hyperbolic theories have problems of their own. The hyperbola is defined by two perpendicular asymptotes, in this case one at the baseline of zero reward value and the other at the time of reward. Inevitably, a hyperbolic theory must predict reward value approaching infinity as the function moves toward the moment of reward. Rewards should become irresistible, no matter what other conditions apply, when the moment of reward gets very close. The exponential curve, in contrast, always has a finite value at a given time. True, the hyperbola has a predictive advantage over most of its range, where it predicts lower reward value (more discounting of future rewards) than exponential models, but the choice of either conic section is more a matter of mathematical convenience than theory, because no reasonably developed theory of reward motivates either model.

Another problem with hyperbolic theories became clear to me when I applied the only hyperbolic model that I have developed during my career, an ideal-observer model that predicts reading rate at any rate of display flicker, as on a CRT monitor. It was a harrowing experience, because once the positions of the two asymptotes have been established (in this case at zero display frequency and reading speed at infinite frequency) there is only one free parameter left, a scaling parameter. It's not much to go on. The model worked well enough, though, enabling us to predict reading speed at one frequency with great precision, given only the reading speed at another frequency (Montegut et al. 1997).

The hyperbolic model of reading rate worked because it dealt with a low-level stimulus sampling issue, not with deep psychological issues of reward and choice. The hyperbola simply doesn't leave enough room to take complexities into account. Recognizing this, Ainslie proposes that reward-delay decisions involve a whole cascade of hidden intermediate decisions, each with its own hyperbolic function that is replaced by another hyperbolic function at the time of an intermediate decision. The resulting predictions of future reward value are neither exponential nor hyperbolic, but depend on the timings of the intermediate decisions. Each intermediate decision brings with it a new free parameter, making the model more predictive but less theoretically useful, because the number of free parameters expands faster than the number of predicted actions.

Ironically, the decision cascade idea highlights an essential paradox of linear decision theories – decisions are themselves nonlinearities, places where everything that has gone before is applied to a single binary choice: you either accept the reward or you

https://doi.org/10.1017/S0140525X0523011X Published online by Cambridge University Press