don't. After each binary choice, represented by the vertical lines in Ainslie's diagrams, the data that previously went into the choice are discarded, and the process starts again. This is the sense in which decisions are nonlinear. And it's difficult to make a linear theory of nonlinear behavior.

Even this expanded theory, however, fails to resolve the deepest paradox of reward theory, the issue of what seems like free will. Even if the intermediate decisions are driven by utility theory or by some other deterministic algorithm, we still do not know how that last decision gets made. Substituting intertemporal bargaining within an individual adds flexibility, but there still must be rules governing each bargaining agent. The paradox of free will remains, the two alternatives of choice determined either by some unknown influence or by uncaused action. In the first alternative, free will is an illusion, a feeling of free choice in an environment where the decision is in fact determined by unconscious information-processing in the brain. The second flies in the face of everything we know about the physical world. Indeterminacy from random or chaotic processes doesn't solve the problem, for it only adds a bit of noise to the reasoning process, whether conscious or unconscious. And noise is not the same as free will; even in introspection, we don't confuse lack of control with freedom. The only alternative left is the uncomfortable first alternative, that is, that free will is indeed an illusion, but since it is a consistent illusion, it is accepted as reality.

The idea that consistent illusions are perceived as reality has precedents from more prosaic, but better-defined, domains in perceptual research. Illusions can be defined as situations that change upon closer inspection, whereas reality remains the same upon closer inspection. Length illusions, such as the Müller-Lyer arrows-in versus arrows-out figure, for example, can be tested easily by measuring the two lines in question, or by superimposing them for direct comparison. The illusion becomes obvious, but even after years of experience the figures still appear to be of distorted lengths. Other illusions are more difficult to expose as illusions. The slopes of hills, for instance, are grossly overestimated by most people, who will go through their lives believing that the steepest streets in San Francisco are perhaps 45 degrees, when they are actually about 10 degrees. If no one corrects them on this, the 45 degrees is reality for them, with none of the conflict that pertains to illusions when they are exposed. Analogously, if the unconscious information-processing that goes into decision-making is never exposed, people can go through their lives believing that their thought processes are guided by free will, and never be confronted by the paradoxes of uncaused action or hidden determinism. Still, a theory that can help us to predict behavior in serious situations, such as addictions, is a pragmatic step forward.

ACKNOWLEDGMENT

Preparation of this commentary was supported by a Faculty Research Grant from the Academic Senate of the University of California, Santa Cruz.

Regret and the control of temporary preferences

Terry Connolly[a] and Jochen Reb[b]

[a]*Department of Management and Policy, The Eller College, University of Arizona, Tucson, AZ 85721;* [b]*Singapore Management University, Lee Kong Chian School of Business, Singapore 178899.* **connolly@email.arizona.edu jochenreb@smu.edu.sg**

Abstract: Regret is often symptomatic of the defective decisions associated with "temporary preference" problems. It may also help overcome these defects. Outcome regret can modify the relative utilities of different payoffs. Process regret can motivate search for better decision processes or trap-evading strategies. Heightened regret may thus be functional for control of these self-defeating choices.

In Ainslie's taxonomy of "temporary preference" problems, the defining feature of addictions is that "the imminent prospect of such activities is strongly rewarding but they're avoided if foreseen from a distance *and regretted afterward*" (Ainslie 2001, p. 49; italics added). In compulsions, similarly, "regret may still occur . . . and the person may even expect the regret while indulging in the behavior" (p. 50), but the behavior still persists. Regret, then, may accompany, or even define, the problematic behaviors. What Ainslie appears to overlook is the possibility that regret may help to control them.

The "temporary preference" problem involves, minimally, the integration of two payoffs, one of which arrives earlier than the other, and is thus prone to the immediacy or hyperbolic discount effect. Viewed from a suitable distance before or after the choice, the later option is superior. Close to the decision point the earlier option is (temporarily) more attractive, a phenomenon that, as Ainslie notes (p. 198), "can't be changed by insight per se." The addict surrenders to the overwhelming desire for a fix, and the compulsive for yet another hand-wash, despite a clear intellectual understanding that she would, on balance, prefer not to and that the temporary urge will pass. Intellection is overpowered by emotion, System 2 by System 1 (Kahneman 2003).

Although emotions commonly involve both cognition and feeling (Frijda 1988), regret does so to an unusually large extent (Landman 1993). Asked if we feel regret over our choice of job, spouse, or vacation, most of us would reply "Let me *think*." We comfort a friend torn by feelings of regret by offering consoling *thoughts*: "How could you have known?" "You made a careful choice, there's nothing to blame yourself for." This intimate interweaving of thought and feeling has made regret the variable of choice for decision theorists interested in emotions. Perhaps it has a role to play in the complete understanding of temporary preference problems.

It is useful to distinguish two targets of decision-related regret: (1) regret associated with the outcome of a decision, and (2) regret associated with the choice process itself (Connolly & Zeelenberg 2002). The former seems to be essentially a reference-point phenomenon, in which the value of an outcome is reduced (or, if positive, enhanced: e.g., rejoicing) by comparison with some reference point, commonly the outcome of a foregone alternative (see Bell 1982; Loomes & Sugden 1982; Mellers et al. 1999). The second seems to involve a stronger component of self-blame or remorse, and is tied to the feeling that the decision made or the process used in making it was, in retrospect, insufficiently justified. The two regrets may be compounded, as for a mother who feels both outcome regret at the sickness of her small child, and self-blame regret at not having thought more carefully about his medical care (Reb & Connolly 2005). The failed dieter regrets both the additional weight gained and the poor decision about the chocolate cake.

There is abundant evidence that anticipated regret can influence decisions in a variety of domains, including medical care (Connolly & Reb 2003), consumer decisions (Simonson 1992), and negotiations (Larrick & Boles 1995). Richard et al. (1996) report some success in one temporary preference context, curbing unsafe sexual behavior, by shifting time-frames and making regret salient. They asked their respondents about either their feelings *about* unsafe sex or the feelings they would anticipate *after having had* unsafe sex. Participants in the second condition reported "safer" behavioral expectations immediately, and less actual risky sexual behavior in the six months following the experiment. How, exactly, might such a manipulation of regret salience achieve this promising result?

Two mechanisms might be suggested, paralleling the two sorts of regret described above. One possibility is that the manipulation enhanced outcome regret associated with the smaller, sooner behavior (unsafe sex), lowering its payoff value, and/or increased the larger, later payoff value by adding a component of rejoicing. It is clear from the conventional portrayal of discounted payoff values (e.g., Ainslie 2001, Fig. 4, p. 63) that modestly lowering the ear-

lier payoff or raising the later one could resolve the preference reversal in favor of the later behavior.

A second, perhaps complementary, mechanism might rely on the self-blame, process-oriented component of regret. We have shown in recent work (Reb 2005) that making regret salient to experimental participants can lead them to use more careful decision processes, acquire more decision-relevant information, and deliberate longer before deciding. Perhaps the participants in Richard et al.'s study responded to the regret-salience manipulation by searching more diligently for alternative choices, weighing the costs and benefits of the unsafe behavior more carefully, or considering one of the familiar self-control strategies discussed by Ainslie (p. 73ff).

The hypothesis, then, is that regret can be more than a mere symptom of failed decision making. Regret may, in some circumstances, play a role in improving decisions: the experience of regret can drive learning in repeated decisions; its anticipation can shape single decisions. Outcome regret affects decisions by modifying the relative attractiveness of different payoffs. Process-related regret does so by motivating the search for trap-evading strategies such as decision bundling, precommitment, and the like. In both cases, the interweaving of thought and feeling that characterize regret provide the bridge between System 1 and System 2 processes, between the thoughtful appraisal of the distant goal and the visceral appeal of the immediate indulgence. Without venturing into evolutionary speculation that regret may have developed to serve such a system-bridging purpose, it is not difficult to see that some level of regret can be highly functional for control of the self-defeating processes that temporary preference problems represent. The hypothesis seems to us worthy of serious consideration.

The will: Interpersonal bargaining versus intrapersonal prediction

Luca Ferrero

Department of Philosophy, University of Wisconsin–Milwaukee, Milwaukee, WI 53201-0413. **ferrero@uwm.edu** **http://www.uwm.edu/~ferrero**

Abstract: Ainslie is correct in arguing that the force of commitments partly depends on the predictive role of present action, but this claim can be supported independently of the analogy with interpersonal bargaining. No matter whether we conceive of the parties involved in the bargaining as interests or transient selves, the picture of the will as a competitive interaction among these parties is unconvincing.

I am unpersuaded by Ainslie's central claim that the will is the product of transtemporal bargaining among successively dominant, transient interests analogous to the emergence of cooperation in a repeated Prisoner's Dilemma (Ainslie 2001, pp. 90–93). It is questionable that we could make sense of the parties involved in this bargaining as truly separate sources of agency. And even if we could, it is hard to see how cooperation could emerge out of the interactions between these parties. It is doubtful, therefore, that the will can be understood as a genuine *inter*personal phenomenon.

Consider the alleged competition among separate interests. Ainslie presents the interests as independent agencies that strive for selection (Ainslie 2001, pp. 39–41, 61, 73) as if they were replicators in a process of natural selection. But he gives no reason to believe that there are heritability and differential fitness in the competition among interests. What the selection amounts to is just that the strongest interest is satisfied at the expense of the conflicting, weaker interests. This satisfaction does not alter the chances of any interest to reappear with equal strength in the future. Nor does it promote the development of any adaptive strategy by the interests themselves. Understanding of the effects of hyperbolic discounting does not depend on the unwarranted reification of the agent's preferences into independent sources of agency that compete strategically in a genuinely selective process. Talk of selection among the preferences and the development of strategies to deal with the conflicts of preferences is more appropriate at the level of agents, even if the agents operate in response to the varying strengths of their preferences. In any event, interaction among interests could not explain the emergence of commitments. A short-range interest has no incentive to submit to a commitment, because commitments preclude the interest's present and future satisfaction. Interests seek nothing other than their satisfaction, hence nothing can be offered to them in exchange for their frustration.

What if the parties are not interests, but successive temporal selves? Ainslie occasionally shifts from talk of transient interests to talk of temporal selves (Ainslie 2001, pp. 40, 93, 161). The two notions are not identical, however. Contrary to transient interests, temporal selves are sources of agency and can have multiple interests. It seems that temporal selves might agree to be under commitments that frustrate their dominant short-range preferences in exchange for the satisfaction of other preferences. However, temporal selves are transient, hence they have no incentive to settle for less than the satisfaction of their short-range dominant interest.

The problem would not arise for parties that are transient in the sense that they act just once, but have stakes in the long-term outcomes of their actions (see Ainslie 2001, p. 93). These parties have no problem seeing the long-term benefits of a commitment. Nevertheless, they are tempted to make an *exception* now, thereby satisfying their dominant short-range interest while still reaping the long-term benefits of future compliance. However, if the present action counts as a precedent, a single exception to the commitment is self-defeating, given that present defections invite future ones. For Ainslie, transient interests/selves happen to be related so that their actions count as precedent for future ones, whence the stability of commitment. However, the fact that transient parties with long-term stakes can strategically agree to cooperate does not explain the will. First, there is no need to look at interpersonal scenarios to appreciate that actions can work as *intra*personal precedents. Given that the *same* agent is going to face the *same* choice at the future time with the *same* set of preferences, it is not surprising that her present action is a precedent for her future ones, thereby defeating temptations to make exceptions to her commitments. This is not really *strategic* thinking, but just reflection on the import of one's present action in the context of one's continued existence as one and the same agent who is going to face exactly similar choices in the future. Appeal to transient selves adds nothing to this straightforward *intra*personal explanation. Moreover, in order to make the repeated Prisoner's Dilemma scenario envisaged by Ainslie (2001, p. 93) truly explanatory, special interpersonal conditions must be assumed: The parties must face exactly the same choice over time and share the same long-term preferences. But these conditions are not special from the intrapersonal point of view. They are just distinctive features of the agent's temporal identity.

Second, the fact that we are subject to hyperbolic discounting and thus prone to inconsistent shifts in short-term preferences is no reason to think that we are made up of competing transient selves with long-term stakes. What makes this false picture attractive is the misleading focus on scenarios like "Ulysses and the sirens" as if they were paradigmatic of diachronic agency. Ulysses' situation, however, is unusual. When Ulysses listens to the sirens, he does not just reverse his short-range preferences, rather, he is also insensitive to long-term considerations. Hence, he does not care that his action could be a precedent. But this makes him impervious to commitments. He can be controlled only by physical restraints or short-range disincentives. If hyperbolic discounting were to make us always like Ulysses, our lives would indeed be best described in interpersonal terms. But then there would be no will, just crude transtemporal manipulation. On the other hand, if temporal selves are depicted as having not just shifting short-term