4

Commitment Devices under Self-Control Problems: An Overview

ISABELLE BROCAS, JUAN D. CARRILLO, AND
MATHIAS DEWATRIPONT

In everyday decision-making, most choices made by individuals involve intertemporal tradeoffs between immediate and delayed gratification (eating, drinking, and smoking versus abstinence to keep a healthy state, current consumption versus saving for the future, leisure versus work to increase wealth, etc.). Traditional economic analyses capture the temporal dimension of individual preferences with the exponential discount utility. This representation of preferences has (among others) two basic properties. First and other things being equal, future utility is valued less than current utility. Second, decisions are consistent over time: The choice between one type of gratification at date t and another one at date $t+s$ depends on the time interval between these two periods (s) but not on the distance between the current period and the first relevant date (t).

Research first in Psychology and now in Economics suggests that individual choices consistently satisfy the first property, except in rare occasions (see, for example, Prelec and Loewenstein 1991). By contrast, it also indicates that they often violate the second property. For example, an individual may prefer a 7-day trip to Europe now rather than a 10-day trip in 1 month and, at the same time, a 10-day trip in 13 months rather than a 7-day trip in 1 year. In this example, the time interval between the two choices (1 month) does not provide enough information to explain which alternative is preferred.

This violation has important consequences: Optimal contingent plans from the current perspective may not be followed if choices can be reconsidered as time passes. For example, despite his current preference for the 10-day trip in 13 months, the individual anticipates that, one year from now, he will not want to wait an extra month to enjoy the longer trip. Strotz (1956) was the first author to formalize this dynamically inconsistency of preferences. The paper proposes a specific direction for the time preference-reversal: The discount rate between two consecutive periods increases as these dates approach. In other words, date 2 is more heavily discounted with respect to date 1 than date 3 with respect to date 2, and so on. Subsequent experimental and empirical research has refined this finding and concluded that the best approximation of time-discounting is

captured with a hyperbolic function. Formally, from the perspective of date 0, date t is discounted at a rate $\phi(t) = (1 + \alpha t)^{-\beta/\alpha}$ (Loewenstein and Prelec 1992). As one can easily check, the marginal rate of substitution between one unit of consumption at date t and one unit at date $t+1$ increases with t (instead of being constant as in the traditional exponential discounting case).

In recent years, this form of time-inconsistent discounting has received much attention from economists. For analytical tractability, many authors have employed the "quasi-hyperbolic" function first introduced by Phelps and Pollak (1968) in the context of imperfect inter-generational altruism: From the perspective of date 0, the individual discounts date t (≥ 1) at a rate $\beta\delta^t$, with $\beta \in (0, 1)$. This reduced-form version loses some of the properties of the pure hyperbolic discounting (in particular, the marginal rate of substitution between consumption at two future consecutive periods is constant). However, it captures its essence: Optimal current plans for future periods are no longer optimal when the time to implement them arrives. Under this approach, the individual is modeled as a collection of independent "selves" with different intertemporal utilities who play a non-cooperative intrapersonal game. The (natural) restriction of the game is the predetermined order of moves: self-1 (i.e. the agent whose intertemporal utility is computed from his date 1 perspective) moves before self-2, who moves before self-3, and so on. It is then solved using the standard Subgame Perfect equilibrium concept. Laibson (1997) was the first author to point out the analogy between this intra-personal game and the imperfect intergenerational altruism model of Phelps and Pollak. Although, this tradition has been pursued by most of the subsequent literature, some authors have proposed other approaches both to model time-inconsistent preferences and to solve the game (see Caillaud and Jullien 2000, for a review). Also, it is fair to say that there is still a substantial amount of controversy on whether hyperbolic discounting is indeed the best way to approximate the dynamic preferences of individuals.[1]

The purpose of this chapter is not to provide a historical perspective of theoretical, empirical, or experimental research on time-discounting. For the readers interested in the history and developments on discounting, we refer to the nice collection of essays in Loewenstein and Elster (1992). For a comprehensive summary of experimental research, see Frederick, Loewenstein, and O'Donoghue (2001), and for a non-technical introduction to hyperbolic preferences, see Ainslie (1992). Our goal is different. First, we review some of the main commitment devices used to increase welfare by individuals with self-control problems (Section 1). Second, we study in more detail the role and credibility of a very natural type of commitment device: A promise. We analyze which type of promises are likely to be made and whether they are honored in equilibrium (Section 2). Last, we draw a parallel between two different literatures: Time-inconsistency and contract incompleteness. In both cases, the individual starts with a commitment problem (with future incarnations of himself and with other

AQ: The reference of Frederick et al. reads 2002 in reference list. please check.

[1] See, for example, the criticisms of Mulligan (1996), Rubinstein (2000), and Read (2001).

individuals, respectively) and uses the available commitment devices to tie up his hands and improve his expected payoff (Section 3).

Before reviewing the different commitment devices, we should note upfront the existence of a conceptual problem when considering welfare issues. Given the tension between the desires of an individual at different dates caused by the dynamic inconsistency of his preferences, some commitment devices may increase the intertemporal welfare from the perspective of some selves but decrease it from the perspective of some others. If this is the case, what is the appropriate way to determine a welfare improvement? Ideally, one should consider a Pareto criterion: The welfare of all selves must be at least equal and that of some selves higher under the new measure. Unfortunately, this criterion is usually hard to meet. Alternatively, O'Donoghue and Rabin (2000) advocate the idea of focusing on welfare from the individual's self-0 perspective, that is before he has to make any decision. Under the quasi-hyperbolic assumption, this amounts to setting $\beta = 1$ when doing the welfare computations. However, there is no normative foundation for this proposal. All these considerations are relevant when we discuss the optimality of policies implemented by a social planner. In most of the chapter, we will ignore this (otherwise important) issue and simply discuss the choices of an individual who maximizes the intertemporal utility from his current perspective given the available commitment devices.

1. SOME COMMITMENT DEVICES

1.1. Interpersonal Commitment Devices

The simplest types of commitment devices employed by individuals to increase their welfare are provided by the market. Laibson (1997) studies dynamic consumption by a hyperbolic discounting agent. Not surprisingly, in the absence of commitment possibilities, the savings rate in future periods is excessively low from the agent's current viewpoint. As a result, illiquid assets become an attractive form of investment, for which the agent may be willing to pay a premium (see also, Harris and Laibson 2003). These investments may result in Pareto superior outcomes. The current self must necessarily benefit from such investment (otherwise he would not incur it). More interestingly, subsequent selves may find that the loss of a reduced freedom can be offset by the gain of higher inherited savings. Thus, the paper highlights the desirability of mandatory pension plans with limited accessibility. Diamond and Koszegi (2000) argue that some of Laibson's conclusions can be modified when the date of retirement is also a choice variable. The paper shows that future selves not only save less than optimal from the current perspective, but also have incentives to retire sooner due to their higher marginal cost of staying active. Anticipating this, earlier selves may decide either to save more to compensate for the (unavoidable) early retirement or to save less so as to "force" future incarnations to stay active. Hence, a mandatory minimum number of years of work before retirement can

act as a commitment device against inefficient saving and retirement decisions. Barro (1999) studies a neoclassical growth model using a continuous-time version of Laibson's hyperbolic discounting setting. This formalization allows the author to analyze the effect of partial commitment, defined as the ability to select the flow of consumption for a given length of time. The comparative statics show that, as the commitment capacity increases, the level of savings increases in the long run (households can better combat their future self-control problem) but decreases in the short run (households can satiate immediate gratification without jeopardizing long-run welfare).

In a different vein, Nocke and Peitz (2001) compare the purchasing decision of hyperbolic discounting agents and the equilibrium prices for durable goods with and without competitive secondary markets. The key issue is that purchasing the good is a commitment for the individual to consume it at every period only if there is no tradable market. Thus, some agents may prefer consumption at every period rather than no consumption ever and yet follow the second alternative because they anticipate that, due to their intrapersonal conflict, a purchase will be inevitably followed by a resale.

Naturally, firms with local monopoly power can exploit the self-control problem of consumers. Wertenbroch (1998) claims that firms should distinguish between two types of goods when exercising second-degree price discrimination. Purchase of "virtue" goods (i.e. goods with low immediate pleasure but positive long-run effects on health like reduced fat, caffeine-free, and alcohol-free products) can be greatly encouraged with minor quantity discounts. By contrast, in order to exercise some self-control, consumers of "vice" goods (i.e. goods with high immediate pleasure but negative long-run effects on health like regular products, alcoholic beverages, and tobacco) are willing to pay substantial premium prices for reduced quantities. In a related spirit, Carrillo (2002) shows that a hyperbolic discounting agent with imperfect knowledge about his taste for quality (flying business, driving expensive cars, etc.) and who learns it through consumption may optimally decide never to try the luxury good. The reason is that if the consumer likes the luxury good only moderately, he will be trapped in an inefficient intertemporal behavior: Each self will consume high quality and sacrifice future wealth. As a reaction, the firm offering these products should try to hook the consumer into high quality. Thus, according to this theory, an airline company should find it more profitable to offer business class upgrades rather than free economy tickets as frequent flyer awards. Similarly, introductory pricing and free test before purchase should be relatively more frequent for goods where the agent learns rapidly his taste for quality than for goods where learning is more gradual.

Markets can affect individuals in other negative ways. Jovanovic and Stolyarov (2000) analyze the effect of technological adoption. The paper argues that hyperbolic discounting agents may prefer to reject modern technologies as a commitment against a level of future work that is judged excessively high from the current perspective. However, as soon as some agents decide to adopt

the more advanced technology, the marginal benefit of labor is raised for all individuals, and therefore every agent (including those who kept the old technology) is tempted to overwork. In other words, market forces destroy the commitment ability and adversely affect the welfare of every agent in the economy. According to the authors, this could explain the health decline in the United States and Britain during the period of industrialization.

The other natural, external commitment devices used by individuals are explicit or implicit contractual relations with third parties. These can take numerous forms, depending on the situation at stake. O'Donoghue and Rabin (1999) study the optimal contract offered by an exponential discounting principal to a hyperbolic discounting agent with private information on the day-to-day cost of undertaking a task. The authors prove the optimality of a "deadline-type" contract, with mild (resp. severe) punishments for small (resp. large) delays before the completion of the task. When the agent is (partially or fully) unaware of his time-inconsistent preferences, then the principal exploits the agent's bounded rationality by offering contracts that specify lenient per-period punishments as long as tasks remain unfulfilled. Such penalties lull the agent into severe and costly procrastination. Ariely and Wertenbroch (2002) argue that deadlines are often self-imposed although, in order to be effective, they need to be enforced in one way or another by third parties. The paper presents the following experimental study. At the beginning of the semester, some students freely choose the date by which they want to hand in each of the three papers that are part of the course requirement, whereas others have externally imposed deadlines. Since, early completion of the paper does not give extra credit and failure to comply with the deadline is costly, students have no external reason for imposing an early deadline on themselves. The results provide compelling evidence that individuals are aware of their self-control problem, since most students choose to set a deadline. However, it also shows that external deadlines act better than self-imposed deadlines as commitment devices to improve performance.

Brocas and Carrillo (2001) study interpersonal competition and cooperation between hyperbolic discounting agents. In this situation, each individual confronts not only future incarnations of his own self, but also current and future incarnations of other individuals. The paper demonstrates that competition for the completion of a project involving an immediate cost and a delayed benefit decreases the agents' incentives to procrastinate (i.e. to undertake the task "too late"), since delays are translated into a lower probability of being first. Similarly, it also decreases the incentives to rush (i.e. to undertake the task "too early") when benefits are immediate and costs are delayed. Therefore, in both cases, competition can be welfare-enhancing for each and every individual.

A last and obvious way to try and get around one's time-inconsistent behavior is to resort to "promises," roughly defined as an implicit contractual agreement with other parties (Carrillo and Dewatripont 2001). What is the cost of breaking a promise? Are self-imposed promises always honored or is it optimal to

make a promise knowing that it will be sometimes broken? If this behavior is anticipated by everyone, does that make the promise ineffective in the first place? This and other questions are explored in Section 2.

Before moving to the analysis of intrapersonal commitment devices, two remarks are in order. First, it is interesting to notice that, in the traditional literature, interactions between agents with conflicting goals are at best neutral and typically a source of inefficiencies. This remains true most of the time under hyperbolic discounting, as witnessed by the previously reviewed papers. However, Brocas and Carrillo (2001) show that the opposite can also be true: Each agent may use competition with the rival as a commitment device against rush or procrastination and, as a result, interpersonal interactions may turn out to be welfare-enhancing. Second, in most of the literature, one of the parties is either time-consistent or acts during at most two periods (making irrelevant the type of discounting). When all individuals are time-inconsistent, then they interact only through the market. In our view, it would be desirable to provide more studies of interpersonal interactions between a restricted number of hyperbolic discounting subjects.

1.2. Intrapersonal Commitment Devices

Individuals may not need the collaboration of other parties in order to self-regulate their behavior. Strategic ignorance can be a powerful self-restraining mechanism. For an individual with time-varying preferences, free information (about his own preferences, the environment, the state of the economy, the opportunities available, etc.) has both benefits and costs. On the one hand, extra news improve the quality of current choices. On the other hand, every piece of information is automatically shared with future incarnations of the agent, who use them optimally from their own perspective. When the expected costs of more suboptimal future choices outweigh the expected benefits of more accurate current decisions, the individual may optimally decide to forego information. Thus, the theory explains why a researcher with meager but encouraging information about the prospects of his research may refrain from acquiring further evidence and undertake his project: Better information can be helpful for his decision but it can also cast some doubts about its quality and lead to beliefs involving perpetual procrastination (Carrillo and Mariotti 2000). For the reader interested in a detailed exposition of these arguments together with an exhaustive analysis of possible applications, we refer to Brocas and Carrillo (2003).

Rational ignorance can also have market consequences. Under time-inconsistency, the strategic ignorance argument implies that cash constrained entrepreneurs find it individually optimal to avoid collecting information about their project before applying for a loan. Banks anticipate this behavior and increase the lending interest rates to compensate for the lack of self-selection among potential entrepreneurs. Overall, individuals are endogenously linked to each other by the dynamic inconsistency of their preferences: The decision to

avoid information is individually rational but makes everyone worse-off through the effect of market interest rates (Brocas and Carrillo 1999).

This model of hyperbolic discounting with imperfect self-knowledge has been extended in two directions, providing new insights about other self-regulation and self-restraining mechanisms. First, Benabou and Tirole (2001) incorporate imperfect willpower, defined as a state-dependent and imperfectly known degree of time-inconsistency. Under imperfect willpower, current conduct affects directly intertemporal welfare but, more importantly, also indirectly because it signals to the agent his own capacity to resist future temptations. The paper argues that under- and over-regulation are two sides of the same coin. As it is already well known, a hyperbolic discounting agent may excessively indulge his vices. More interestingly, if that same agent has a high level of insecurity about his willpower, then he may adopt an excessively rigid conduct for the sole purpose of reassuring himself about the strength of his character. Second, Benabou and Tirole (2002) consider an individual with imperfect memory about his past behavior. Naturally, the individual would like to remember information that fosters his motivation and forget information that undermines it. Unfortunately, he cannot selectively decide which information is retained and which one is not. As a result, multiple modes of cognitive behavior are self-sustainable: Under high (low) self-censoring of information a lack of news is more (less) strongly interpreted as bad news, which makes the agent more (less) prone to engage in self-repression in a first place. Some other findings of the imperfect willpower and imperfect recall theories are reviewed in Benabou and Tirole (2003).

Last but not least, Caillaud, Cohen, and Jullien (1996) propose a radically different approach to the agent's intrapersonal conflict of preferences. Instead of modeling the individual as a collection of selfish incarnations who play a non-cooperative game "against" each other, the authors consider the individual as one entity who chooses a plan of actions (as opposed to just a current action) on which he can expect to conform not only today but also in the future. Under this approach, the individual disregards inconsistent plans of the type "indulge the vice only today," as they would be revised again and again. The paper shows that the resulting self-restrained strategy is unique, internally consistent (i.e. the individual never has an incentive to change his plan of actions) and characterized by some degree of self-regulation.

2. PROMISES AS A NATURAL COMMITMENT DEVICE

As mentioned in the previous section, promises are probably the most basic form of commitment with third parties. We all make promises to our spouse, to our friends, and to our boss. We promise that we will quit smoking, that we will be on time for the appointments, and that we will complete projects by their due date. However, one may wonder how effective and how credible these promises are. In fact, the two questions are closely related in some kind of paradoxical way.

On the one hand, third parties believe a promise only if the person who makes it incurs a cost of not honoring it. On the other hand, for such a cost to exist it is necessary that third parties believe in the promise. So, for example, if my suppliers promise to deliver the merchandise by a certain date and I take some measures because I anticipate they will be delayed and bring it instead one week later, then there is no real cost for them to break the promise. This makes it ineffective and not credible in a first place. In some academic institutions, punctuality in seminars is another good case of commitments (or promises) that are not costly to break because nobody believes them in a first place.

The purpose of this section is threefold. We start by assuming the existence of an exogenous cost of breaking promises. Given this cost, we illustrate with a simple example the circumstances under which a hyperbolic discounting agent has incentives (*a*) to avoid making promises, (*b*) to make promises that will be fully honored, and (*c*) to make promises that will be partly broken (Section 2.1). Also, we discuss the role of self-imposed deadlines as a particular type of promises and analyze the relative merits of flexible versus rigid deadlines (Section 2.2). Last, we present some microeconomic foundations for the effectiveness of promises. We discuss how promises can be credible *vis-à-vis* of third parties and, at the same time, partly broken in equilibrium (Section 2.3).

2.1. A Simple Illustration of the Efficient Use of Promises

Consider the following stylized example (this section is a simplified version of Section 2 in Carrillo and Dewatripont 2001). An individual lives three periods, $t \in \{0, 1, 2\}$. In period $t = 1$, he decides whether to perform 0, 1, or 2 tasks. Each task has an immediate cost of effort e and a delayed return normalized to 1 and enjoyed at $t = 2$. We assume that the agent has a hyperbolic discounting utility function with $\beta < 1$ and, for simplicity, $\delta = 1$. It means that, from the current perspective, all future dates are discounted at the same rate β. We assume that the cost of effort is $e \in (\beta, 1)$. This provides the strongest conflict of preferences between the desires of the agent from his date-0 and his date-1 perspective. For the former, the order of preferences from best to worst option are to perform 2 tasks, 1 task, and 0 tasks (formally, $\beta(2 - 2e) > \beta(1 - e) > 0$). For the latter, the order of preferences from best to worst are to perform 0 tasks, 1 task, and 2 tasks (formally, $0 > \beta - e > 2\beta - 2e$). So, in the absence of promises or some other commitment device, self-0 would like to perform two tasks at date 1, but is aware that self-1 will not perform any.

Suppose now that self-0 can make a promise to a third party about the number of tasks to be performed by self-1. Making a promise is costless as long as it is fulfilled or surpassed. By contrast, breaking a promise is costly. We assume that the individual cannot credibly promise to perform more than two tasks. Failing short of one task (i.e. promising 2 tasks and performing 1 or promising 1 task and performing 0, from now on denoted [0/1] and [1/2]) has a cost c_L (> 0). Failing short of 2 tasks (i.e. promising 2 tasks and performing 0,

from now on denoted [0/2]) has a cost c_H ($>c_L$). The following table summarizes self-1's payoff for every possible promise made by self-0 and every possible effort realized by self-1.

Self-0	Self-1		
	Effort 0	**Effort 1**	**Effort 2**
Promise 0	0	$\beta - e$	$2\beta - 2e$
Promise 1	$-c_L$	$\beta - e$	$2\beta - 2e$
Promise 2	$-c_H$	$\beta - e - c_L$	$2\beta - 2e$

Self-1's optimal behavior will crucially depend on the costs of breaking promises c_L and c_H. We can distinguish between four cases.

Case 1: $c_L < e - \beta$ and $c_H - c_L < e - \beta$. When both costs are sufficiently small, it is never profitable to sustain a promise. The behavior of self-1 is then [0/1] and [0/2]. Anticipating that promises are useless to elicit effort, self-0 optimally decides not to make any promise. In equilibrium, the agent performs 0 tasks.

Case 2: $c_L < e - \beta$ and $c_H - c_L > e - \beta$. When the cost of failing short of one promise is small and the cost of failing short of two promises is large, then for any level of promise, self-1 falls short of exactly 1 task ([0/1] and [1/2]). Clearly, self-0 prefers [1/2] to [0/1]. The remaining question is whether two promises and only one fulfilled is better than no promise and no task. From self-0's perspective, this is the case if and only if $\beta(1 - e - c_L) > 0$.

Case 3: $c_L > e - \beta$ and $c_H < 2(e - \beta)$. When the cost of failing short of one promise is large and the marginal cost of failing short of the second promise is small, then one promise is maintained but two promises are broken ([1/1] and [0/2]). Self-0, thus, prefers a mild commitment honored rather than a strong one broken. In equilibrium, self-0 makes one promise and self-1 performs 1 task.

Case 4: $c_L > e - \beta$ and $c_H > 2(e - \beta)$. When both costs are sufficiently large, promises are always fully honored ([1/1] and [2/2]). In equilibrium, self-0 promises 2 tasks and self-1 performs both of them.

Overall, depending on the cost structure we may observe no promise, fully honored promise(s), or partly broken promises. The last case is probably the most surprising one: The individual makes a promise anticipating that it will be partly useful as a commitment device and partly not. For this to occur, the marginal cost of breaking the second promise must necessarily be higher than the marginal cost of breaking the first one ($c_H > 2c_L$, see case 2). Conversely, a necessary condition to observe that self-0 moderates his promises and then fully honors them is that the marginal cost of breaking the first promise be higher than the marginal cost of breaking the second one ($c_H < 2c_L$, see case 3).

2.2. Deadlines: Another Example of Promises

A self-imposed deadline is a form of promise that a hyperbolic discounting individual might use to avoid delaying unpleasant tasks (see Ariely and Wertenbroch 2002). However, the optimal "strength" or "tightness" of a deadline must balance commitment and flexibility. On the one hand, a strict deadline is most effective in avoiding procrastination, just like a strict rule of behavior (no smoking) is a more adequate self-disciplining device than a lenient one (smoking "a few" cigarettes). On the other hand, a flexible rule allows the individual to deal better with unforeseen contingencies and to allocate time, effort, and attention more efficiently as new alternatives become available.

The relative advantages of rigid versus flexible deadlines can be illustrated with the following stylized example which, as usual, does not pretend to capture the full complexity of real life situations. Consider an individual with a stochastic probability of being offered short-term jobs. For simplicity, we assume that in period $t=1$ he is offered 1 job with probability $1-p$ and 2 jobs with probability p and in period $t=3$, he is offered 0 jobs with probability $1-q$ and 1 job with probability q. He can perform at most one job per period and only in periods 1, 2, and 3. The job requires an immediate effort which has a cost e and yields a one-period delayed return normalized to 1. As before, the individual has a hyperbolic discounting utility with $\beta<1$ and $\delta=1$. Although jobs have positive net present value (i.e. $-e+\beta>0$), the individual has incentives to procrastinate. More precisely, if one job is pending at $t=2$, the individual prefers to wait until $t=3$ to perform it, even at the risk of not being able to accept another job that could be offered at that date (formally, $\beta[-e+1]>-e+\beta[1+q(-e+1)]$). Note that both conditions are compatible when the cost of effort is $e\in(\beta q/(1-\beta+\beta q),\beta)$. Therefore, time-inconsistent preferences ($\beta<1$) and a positive likelihood of no job being offered at $t=3$ ($q<1$) are both necessary conditions. Under these circumstances, one can immediately see that if the individual is offered one job at $t=1$, then he will procrastinate until $t=3$. If he is offered two jobs at $t=1$, then he will perform one at $t=2$ and one at $t=3$.

Can self-imposed deadlines improve performance? Consider first the most rigid one: "Either I do the job the date in which it is offered or I refuse it." From the perspective of self-0, this type of deadline is preferable to no deadline if and only if

$$\beta[-e+1+q(-e+1)]>\beta[p(-e+1)+(-e+1)]\Leftrightarrow q>p,$$

where the l.h.s. is the expected payoff under a rigid deadline (one job performed at $t=1$ and another one at $t=3$ if it is offered) and the r.h.s. is the expected payoff under no deadline (one job at $t=3$ and another one at $t=2$ only if two jobs are offered at $t=1$). Note that both strategies have costs. Under no deadline, the individual never takes advantage of the job possibly offered at $t=3$ due to his incentives to procrastinate. Under a rigid deadline, the individual must

forego one job if two are offered at $t=1$. Consider now a more flexible rule: "Any job accepted is performed on the date it is offered or on the next one but not later." This more lenient rule combines the benefits of flexibility (the second job offered at $t=1$ is not necessarily refused) and commitment (a three-period delay is not acceptable). In fact, it is easy to see that if one job is offered at $t=1$ then it is performed at $t=2$, and if two jobs are offered at $t=1$ then one is performed at $t=1$ and the other at $t=2$. Also, any job offered at $t=3$ is performed at that date. Thus, in our simple example, this somewhat permissive deadline achieves self-0's first-best outcome and an intertemporal payoff of $\beta(-e+1)(1+q+p)$. Needless to say, the example is somewhat contrived. However, the tradeoff between flexibility and commitment will be present to a greater or lesser extent in any dynamic model with hyperbolic discounting agents and uncertainty.

2.3. Some Microfoundations for the Effectiveness of Promises

The existence of an exogenous cost of breaking promises (Section 2.1) or not meeting deadlines (Section 2.2) helps the individual to increase the production level and the number of tasks undertaken. Yet, it seems that rational expectations and unfulfilled promises are somewhat in contradiction. Suppose that delivering one item is costly only when the principal has adjusted the production process for two. If promising two items automatically indicates that one and only one will be delivered (as, for example, in case 2 of Section 2.1), it should be perfectly anticipated by the rational principal when he adjusts the process, making the promise useless and irrelevant in a first place.

The key issue is to understand what is the real cost of breaking a promise, how it can be anticipated by all parties and yet suffered in equilibrium. There are two ways to approach this question (the arguments provided below are developed more formally in section 3 of Carrillo and Dewatripont 2001). First, suppose that the principal has incomplete information about how costly it is for the agent to exert effort. A low-cost type always promises and delivers high effort. A high-cost type also promises high effort as a commitment device to exert (at least) a medium level of effort. Promises are credible because the principal is unsure about which individual has made them. Also, reneging on a promise is costly as it reveals the agent's type. Second, consider a multi-person contractual relationship. An individual hires a monitor who checks per-period effort. The promise of payment if shirking is detected by the monitor acts as a commitment device. However, imperfect monitoring will result in partial under-provision of effort as the agent anticipates that some shirking can go unnoticed.

To sum up, both the endogenous reputation cost (possible due to incomplete information) and the endogenous financial cost (created with the contractual agreement) can lead to situations where time-inconsistent individuals optimally choose to make promises knowing that these will be broken—partially or with positive probability—in equilibrium.

3. TIME-INCONSISTENT PREFERENCES OR CONTRACT INCOMPLETENESS?

In this final section, we argue that the commitment devices discussed above remain relevant when we replace time-inconsistent preferences by limits to contracting as the source of the commitment problem of the individual. Indeed, what the results rely on is the fact that the individual starts with a commitment problem. As is well-known in economics, such problems can arise in the absence of time-inconsistent preferences: In strategic situations, tying one's hands in advance can be helpful as a way to influence the behavior of others. This is, for example, the case whenever an individual has to contract in the presence of moral hazard (hidden actions) or adverse selection (hidden information) problems, and has to bear a positive share of the resulting inefficiency loss. In this section, we first discuss a variety of well-known commitment problems with time-consistent individual preferences where commitment devices would naturally be useful. Then, we argue that limited information has been stressed as a way to get around commitment problems, in a way that is somehow similar to the literature that has started with Carrillo and Mariotti (2000). Last, following Carrillo and Dewatripont (2001), we detail how we could reinterpret the analysis of promises in terms of limited contracting, and this whether we take the reputational or the explicit monitoring foundations for their effectiveness as exposed in Section 2.3. The connection between time-inconsistent preferences and limits to contracting is a very general point, and one which may not surprise some readers, although we have not found it made elsewhere in the literature. From a general perspective, it means that the literatures on time-inconsistent preferences and on limits to contracting could usefully learn more from one another.

3.1. Commitment Problems with Time-Consistent Preferences

While the focus so far has been on commitment problems arising from hyperbolic discounting, they are also present in a variety of strategic situations with limits to contracting. Let us just mention a few well-known examples.

First, the classical *moral hazard* problem where the agent faces a competitive supply of principals (see Jensen and Meckling 1976 as one example out of a large literature), and where the agent's "effort" is privately determined so that he ends up maximizing his payoff given the contract that has been offered to him. This implies underprovision of effort if the contract offers insurance, or more generally if part of the return from higher effort accrues to the principal (e.g. as a repayment for his initial financing, in a corporate finance context). The agent's inability to commit to exert the effort that would maximize the joint payoff of both parties hurts him in terms of *ex ante* expected utility: The potential principals all ask for the expected market return, in equilibrium they are not fooled, and therefore the agent ends up bearing the cost of his *ex post* distorted effort

choice. He would thus happily take advantage of a commitment device to bring his effort choice closer to the first-best level.

Second, the *ratchet effect*, which arises, for example, in the context of a Government that regulates a firm which has private information about its costs (see, for example, Freixas, Guesnerie, and Tirole 1984; Laffont and Tirole 1988). The ratchet effect is defined as the impossibility for the principal to commit not to take advantage of productive firms by subsequently raising their required output or lowering their cost reimbursement. Since the firm rationally predicts this problem, it reacts by underproviding effort so as to look inefficient, and therefore in need of more lenient production requirements or more generous cost reimbursements. This problem has been emphasized both in the context of the former centrally planned economies and in the regulation of natural monopolies in market economies. It applies potentially in any organization where adverse selection is important and where the principal cannot protect agents with binding long-term contracts.[2]

Third, the *soft budget constraint* syndrome, defined as the inability of investors to terminate loss-making projects, again an issue that has received attention both in the centrally planned and the market contexts. In the presence of sunk costs, refinancing bad projects can be sequentially optimal, because their present value gross of the funding that has already been sunk may be positive even though their net present value is negative. When only the entrepreneurs asking for money know the project returns, investors may start funding projects that have uncertain returns and find optimal to keep funding them even though they feel sorry to have started doing it in the first place. If, however, it were true that the threat of termination (i.e. no refinancing of bad projects) could deter bad entrepreneurs from asking for funds in the first place, investors would be looking for commitment devices not to refinance bad projects (see, for example, Dewatripont and Maskin 1995).

In all these cases, the individual who has a commitment problem can be reinterpreted as having two different "selves": Self-0, before he signs the contract or gets information, and self-1, once the contract has been signed or information has been transmitted. Self-0 would like to commit self-1 in a way that self-1 would like to resist. The parallel between the commitment devices that can be envisaged under individual time-inconsistent preferences and contractual problems with time-consistent preferences is, therefore, quite natural.

3.2. COMMITMENT THROUGH LIMITED INFORMATION ACQUISITION

A common theme in the literature has been the role of limiting the information available to the uninformed party as a way to mitigate his commitment problem.

[2] In fact, a variation on this problem even arises in the presence of long-term contracts when the parties cannot commit not to engage in mutually profitable renegotiations *ex post* (see, Dewatripont 1989).

This arises very naturally in adverse selection problems. First, in the presence of the ratchet effect, Laffont and Tirole (1988) derive conditions under which it is, in fact, impossible to have a contract whose execution would perfectly inform the principal about the agent's type: The agent always prefers to keep looking inefficient. Second, in the soft-budget-constraint case, Dewatripont and Maskin (1995) discuss how decentralizing credit, by creating informational problems (namely, by assuming an informational asymmetry between the initial investor and the refinancing investor), may lower the efficiency of the refinancing activity sufficiently so as to discourage it altogether, thereby deterring bad entrepreneurs from asking for funds at the initial stage.

The usefulness of limited information as a commitment device has also been discussed in moral hazard contexts. Crémer (1995) looks at a model where the agent's performance depends on his (exogenous and initially unknown) talent on top of his effort. The question is whether the principal would benefit from or instead be hurt by his ability to observe the agent's talent at an interim stage. Crémer assumes that the principal could commit not to observe it by keeping the agent "at arm's length" (e.g. as an independent contractor rather than as an employee). The value of doing this is that it would strengthen the agent's effort incentive: By not being able to disentangle effort from talent, the principal would attribute lower output by the agent partly as the result of lower talent, which would reduce his incentive to keep the agent on board. Once again, the benefit of limiting the principal's information arises from the fact that he is unable to commit, this time concerning whether to lay off the agent. Aghion and Tirole (1997) study a problem where the principal wants to induce the agent to acquire costly information about potential project returns. Due to limited congruence between the two parties, the favorite project of the principal is not the same as that of the agent. Knowing that the principal will overrule him discourages the agent and reduces his effort. Aghion and Tirole show how limiting the principal's information about project returns can empower the agent (he acquires real authority) and thereby increase his effort. This benefit can, in fact, outweigh the principal's loss from having the agent's favorite project being chosen rather than his own. Once again, the benefit from limited information arises from the inability of the principal to commit not to overrule the agent.

Note that limiting the information of the principal is the same device used in both the adverse selection and moral hazard contexts. Interestingly, in some situations (Laffont and Tirole 1988; Aghion and Tirole 1997) the goal is not to be too harsh on the agent, while in others (Crémer 1995; Dewatripont and Maskin 1995) the goal is not to be too soft.

We have highlighted the parallel between the limited contracting paradigm and the individual time-inconsistent preferences paradigm (Section 1.2) in stressing the value of ignorance as a commitment device. This parallel has been insufficiently recognized until now and can probably lead to further cross-fertilization.

3.3. Commitment Through Promises

Just as strategic ignorance can be useful under limited contracting, so can promises. In fact, we can reinterpret the setup of Section 2.1 as follows. Instead of thinking of time-inconsistency as the reason behind the fact that incentives to exert effort for the agent are lower at $t=1$ than at $t=0$, one can think of a traditional *hold-up* problem: At the beginning of the game, a principal would be ready to give the agent an amount equal to 1, the return of his effort (which costs e, with $e \in (\beta, 1)$), but could not commit not to renegotiate this to β once effort has been chosen by the agent. This ends up distorting the effort choice of the agent, just as in Section 2.1.

Note that one could think that, initially, the agent could obtain an additional fixed fee from the principal, equal, for example, to $(1-\beta)$ if we assume that the agent has full bargaining power at that initial point and the principal expects to earn this amount later on. However, this will not prevent the effort distortion: Since the agent understands that his marginal return from effort is only β, he cannot commit to exert optimal effort, and thus suffers from exactly the same commitment problem as before.

In this setup, incomplete information and promises can help, exactly as indicated in Section 2.3. Under incomplete information about the agent's cost of effort, there is the opportunity for high-cost agents to make the same promises as low-cost ones as a commitment device to exert more effort. This is discussed in Carrillo and Dewatripont (2001), who also show that the monitoring microfoundation briefly discussed in Section 2.3 above applies in the case of contractual incompleteness too.

REFERENCES

Aghion, P. and Tirole, J. (1997), "Formal and real authority in organizations." *Journal of Political Economy*, 105: 1–29.

Ainslie, G. (1992), *Picoeconomics. The Strategic Interaction of Successive Motivational States within the Person* (Cambridge University Press).

Ariely, D. and Wertenbroch, K. (2002), "Procrastination, deadlines, and performance: Self-control by precommitment." *Psychological Science*, 13(3): 219–24.

Barro, R. J. (1999), "Ramsey meets Laibson in the neoclassical growth model." *Quarterly Journal of Economics*, 114: 1125–52.

Benabou, R. and Tirole, J. (2001), "Willpower and personal rules." Mimeo, IDEI Toulouse and Princeton.

—— and —— (2002), "Self-confidence and personal motivation." *Quarterly Journal of Economics*, 117(3): 871–915.

—— and —— (2003), "Self-knowledge and self-regulation: An economic approach." In I. Brocas and J. D. Carrillo (eds.), *The Psychology of Economic Decisions, Vol.1: Rationality and Well-being* (Oxford University Press).

AQ: Please provide places of publication.

Brocas, I. and Carrillo, J. D. (1999), "Entrepreneurial boldness and excessive investment." *CEPR Discussion Paper 2213*, London.

——— and ——— (2001), "Rush and procrastination under hyperbolic discounting and interdependent activities." *Journal of Risk and Uncertainty*, 22(2): 141–64.

——— and ——— (2003), "Information and self-control." In I. Brocas and J. D. Carrillo (eds.), *The Psychology of Economic Decisions, Vol.1: Rationality and Well-being* (Oxford University Press).

Caillaud, B. and Jullien, B. (2000), "Modeling time-inconsistent preferences." *European Economic Review*, 44: 1116–24.

——— Cohen, D., and Jullien, B. (1996), "Towards a theory of self-restraint." Mimeo, Paris.

Carrillo, J. D. (2002), "To be consumed with moderation." *European Economic Review*, forthcoming.

——— and Dewatripont, M. (2001), "Promises, promises ...." *CEPR Discussion Paper 2680*, London.

——— and Mariotti, T. (2000), "Strategic ignorance as a self-disciplining device." *Review of Economic Studies*, 67: 529–44.

Crémer (1995), "Arm's length relationships." *Quarterly Journal of Economics*, 110: 275–96.

Dewatripont, M. (1989), "Renegotiation and information revelation over time: The case of optimal labor contracts." *Quarterly Journal of Economics*, 104: 589–619.

——— and Maskin E. (1995), "Credit and efficiency in centralized and decentralized economies." *Review of Economic Studies*, 62: 541–55.

Diamond, P. and Koszegi, B. (2000), "Quasi-hyperbolic discounting and retirement." *Journal of Public Economics*, forthcoming.

Frederick, S., Loewenstein, G., and O'Donoghue, T. (2002), "Time discounting and time preference: A critical review." *Journal of Economic Literature*, 40: 351–401.

Freixas, X., Guesnerie, R., and Tirole, J. (1984), "Planning under incomplete information and the ratchet effect." *Review of Economic Studies*, 52: 173–91.

Harris, C. and Laibson, D. (2003), "Hyperbolic discounting and consumption." In M. Dewatripont, L. P. Hansen, and S. Turnovski (eds.), *Advances in Economics and Econometrics—Theory and Applications, Eighth World Congress*, Vol. 1, pp. 258–97 (Cambridge University Press).

Jensen, M. and Meckling, W. (1976), "Theory of the firm: Managerial behavior, agency costs and ownership structure." *Journal of Financial Economics*, 3: 305–60.

Jovanovic, B. and Stolyarov, D. (2000), "Technology adoption and commitments: The welfare costs of Industrialization revisited." Mimeo, NYU and University of Michigan.

Laffont, J. J. and Tirole, J. (1988), "The dynamics of incentive contracts." *Econometrica*, 56: 1153–75.

Laibson, D. I. (1997), "Hyperbolic discount functions, undersaving, and savings policy." Mimeo, Harvard University.

Loewenstein, G. and Elster, J. (1992), *Choice Over Time* (New York: Russell Sage Foundation).

——— and Prelec, D. (1992), "Anomalies in intertemporal choice: Evidence and an interpretation." *Quarterly Journal of Economics*, 107: 573–98.

Mulligan, C. (1996), "A logical economist's argument against hyperbolic discounting." Mimeo, U. Chicago.

Nocke, V. and Peitz, M. (2001), "Hyperbolic discounting and secondary markets." Mimeo, Oxford University.

AQ: Please provide place of publication

AQ: Please update the references of Carillo. J. D. (2002) and Diamond and Koszegi (2000)

AQ: Please provide place of publication

O'Donoghue, T. and Rabin, M. (1999), "Incentives for procrastinators." *Quarterly Journal of Economics*, 114: 769–816.

—— and —— (2000), "Addiction and present-biased preferences." Mimeo, Cornell and Berkeley.

Phelps, E. S. and Pollak, R. A. (1968), "On second best national saving and game-theoretic growth." *Review of Economic Studies*, 35: 185–99.

Prelec, D. and Loewenstein, G. (1991), "Negative time preference." *American Economic Review*, 81: 347–52.

Read, D. (2001), "Is time-discounting hyperbolic or subadditive?" *Journal of Risk and Uncertainty*, 23(1): 5–32.

Rubinstein, A. (2000), "Is it 'Economics and Psychology'?: The case of hyperbolic discounting." Mimeo, Tel Aviv and Princeton.

Strotz, R. H. (1956), "Myopia and inconsistency in dynamic utility maximisation." *Review of Economic Studies*, 23: 166–80.

Wertenbroch, K. (1998), "Consumption self-control by rationing purchase quantity of virtue and vice." *Management Science*, 17(4): 317–37.