CentER

Discussion Paper

No. 2004–49

RELATIONSHIPS AS COMMITMENT DEVICES: STRATEGIC SILENCE

By A. Nir

May 2004

ISSN 0924-7815

TILBURG UNIVERSITY

Relationships as Commitment Devices: Strategic Silence*

Arad Nir?

April 2004

Abstract

People who understand self-control problems can devise mechanisms to overcome them. In this paper, we discuss how relationships can help individuals overcome their self-control problems by creating a tradeoff between desired present procrastination and undesired future procrastination. Threatening not to speak to a person who caves in can create such a tradeoff. The results depend on a limited memory assumption. We show how such interactions can explain strategic pretence, strategic ignorance, why a person would choose to punish himself after he caved in and why punishments need to increase if not adhered to immediately.

JEL *Classification*: D10, D74, D90, R20

Keywords: present-biased preferences, hyperbolic discounting, self-control, commitment device, impulsive behavior, relationships, strategic silence, self-punishment, pretence, emotions, limited memory.

* I thank Chaim Fershtman, Ady Pauzner, Jan Potters, seminar participants at Tel Aviv University and at the EDP Jamboree in London for valuable comments and discussions. This research was partly conducted at Tilburg University while receiving a Marie Curie fellowship. I thank these institutions.

? The Berglas School of Economics, Tel Aviv University, Tel Aviv, Israel. E-mail: aradn@yahoo.com

1 Introduction

Many people have self-control problems. We sometimes engage in activities with immediate rewards (e.g., smoking and drinking) that we would prefer not to engage in the long run, and we sometimes procrastinate indefinitely in performing activities with immediate costs (e.g., revising a paper, exercising, keeping to a diet, saving) that, in the long run, we would have always preferred to perform immediately. People often use commitment devices to deal with these self-control problems. They sign up to give a seminar in order to complete a paper, hold non-liquid assets, or avoid carrying a credit card. Institutions such as Alcoholics Anonymous, drug clinics and workplace rules (office hours are from 9:00 am to 5:00 pm) can also serve as commitment devices.

In this paper, we demonstrate how individuals can overcome their self-control problems when engaged in a relationship without relinquishing discretion of whether to indulge in the problematic activity. This is accomplished by using a strategy that creates a bond between desired present procrastination and undesired future procrastination. The strategies include not speaking to a person who caved in and a self-punishment if he does cave in. We need not assume that not speaking entails disutility per se, rather it's disutility arises because it causes a future self-control problem. The individual punishes himself if he caved in because it facilitates overcoming the self-control problem in the future.

Schelling (1992) and Elster (2000) discuss how individuals can overcome their self-control problems. The tactics they suggest are: relinquishing control to others; disabling (for example: giving away one's credit card, incarceration); imposing penalties and delays on future behavior; removing tempting substances; and quitting precursors

(not drinking because desiring to quit smoking when one often smokes during or prior to drinking). Carrillo and Marriotti (2000) show that people might strategically forgo information in order to commit to a certain course of action. Ainslie (1991) suggests that individuals "bunch" several tasks together in a way that results in a tradeoff between performing a task at present and not performing that task in the future.

> The person's best move at present will thus depend on how he forecasts his future perceptions. Insofar as he sees his current choice as a precedent and not an isolated instance, he will face the incentives of a repeated prisoner's dilemma (Ainslie, p. 337)

Benabou and Tirole (2001) show how individuals can overcome self-control problems by applying personal rules. Battaglini, Benabou and Tirole (2001) show how peer groups affect the individual's ability to overcome self-control problems. Each individual learns about his ability to cope with self-control problems by observing whether others caved in. Benabou and Tirole (2002) develop a Bayesian theory of self-deception with imperfect memory and selective recall. The model highlights the importance of self-confidence in overcoming self-control problems. Our model builds on Ainslie's (1991) idea but within the context of a relationship. That is, we show how a person with a self-control problem and preferences that cause him to repeatedly cave in can overcome his self-control problem by bunching present and future tasks when he is engaged in a relationship.

Consider a couple, Fred and Wilma. We assume that Fred has a drinking self-control problem in the sense that, during each period, he prefers to drink at present but not in the future. Fred drinks at each period and fully understands his self-control problem. If there

were a commitment device available that could force Fred not drink in the future, Fred would be willing to pay for such a device. However, we assume that such a commitment device is not available. Fred cannot sacrifice his discretion, commit to pay a fine in the future if he acts in a certain manner in the present or future, or change any of the fundamental assumptions of the model.

Wilma can help Fred overcome his drinking. She tells Fred not to drink under the threat that if he drinks, she will stop telling him not to drink — that is, she will not talk to him — for some periods. If Wilma's threat is credible, Fred will forgo drinking in the present in order not to drink in the future because during periods when Wilma is not talking to him, he does not have any incentive not to drink. If Fred knows that Wilma will forgive him and continue to talk to him if he drinks, he will continue to drink because the threat is not credible. Speaking or not to Fred is a binary signal that Wilma transmits. When Wilma is talking to him, Fred can influence the future by drinking and causing Wilma not to talk to him in the future. If Wilma is not talking to him in the present he cannot influence the future and drinking is the best action at present. If Wilma is talking to Fred in the present he chooses not to drink because he does not want to enter a state in which Wilma is not talking to him and he drinks. We assume that Fred's memory is limited in the sense that he cannot remember whether he previously drank. The limited memory is necessary so that Fred will not be able to alter the history of future periods by drinking at present.

Now consider the case where Fred is uncertain whether Wilma's preferences are such as to make her threat credible. Under such circumstances, Wilma's threat will be credible because, even if she wants to forgive Fred, she does not want to reveal herself as

a person whose threats are not credible. When Fred is uncertain about Wilma's preferences, he might prefer to remain uncertain about Wilma's preferences: He may not want to learn that Wilma prefers to forgive him because then he will drink in all future periods, an act he wishes to avoid. Such a result is reminiscent of Carrillo and Mariotti's (2000) "strategic ignorance" result, where a person wishes to remain ignorant in order not to bring about a self-control problem. In our model, the agent already has a self-control problem. He wishes to remain ignorant in order for a threat to be credible so that he will be able to overcome the self-control problem.

Last, we show that when Wilma wishes to forgive Fred if he drinks and Fred has full knowledge of her preferences, she can persuade him not to drink by telling him not to drink and specifying a punishment that he must undergo if he does drink. If he does not punish himself, she does not tell him not to drink until he does. The punishment must intensify at a sufficient rate over time so that Fred will not procrastinate punishing himself. Wilma must be indifferent toward the punishment. If she enjoys Fred's being punished, she might blackmail Fred into undergoing it even if he does not drink. If she suffers from Fred's being punished, she will forgive him if he does drink and talk to him without having him enduring the punishment. In the model, even though Wilma must be indifferent toward Fred's punishment, she strictly prefers her actions.

A preference for self-punishment before the person caved in, is straightforward: Punishment is a cost for committing the present biased task; if the cost is sufficiently great, the person will not cave in. After a person has caved in, however, utility from the task is considered a sunk cost. Self-punishment is a cost that people prefer to inflict upon

themselves and thus an economic puzzle. Our model explains why such self-punishments might be rational in some cases.

Commitment mechanisms can be divided into personal and social mechanisms. The social mechanisms literature discusses either surrendering control to others [Schelling (1992), Elster (2000)] or informational externalities [Battagline, Benabou and Tirole (2001)]. The debate about self-control problems between people who are close to one another suggests that the social interaction might be more complex than either mechanism implies. We will thus examine how strategic interactions between couples can solve self-control problems.

We model self-control problems by assuming that people have a taste for immediate gratification and thus place greater weight on present utility relative to future utility. Phelps and Pollak (1968) introduced such present-biased preferences into the study of intergenerational preferences; more recently, Laibson (1994) employed these preferences to model individual intertemporal choice in the context of consumption and savings. O'Donough and Rabin (1999, 2001) have used present-biased preferences to study procrastination[1]. A display of such preferences implies that people might have a self-control problem. Hyperbolic discounting is another term often used for such preferences[2].

[1] Other papers on present-biased preferences include: Akerlof (1991), Laibson (1997), Carrillo and Marriotti (2000), Benabou and Tirole (2001, 2002) and Battagline, Benabou and Tirole (2001) among others. For a review of time discounting and time preferences see Frenderick, Loewenstein and O'Donoghue (2002).

[2] Experimental evidence for hyperbolic discounting was first offered by Thaler (1981) using hypothetical questions and found that the discount rate declines sharply with the length of time to be waited. Benzion, Rapoport, and Yagil (1989) replicated these findings for undergraduate and graduate students of economics and finance at two Israeli universities. Kirby and Herrnstein (1995) demonstrated the same results using real monetary and non-monetary rewards. Kirby (1997) showed that subjects were present-biased using real rewards when they were induced to convey their true value using second-bid auctions.

Strotz (1956) and Pollak (1968) formulated two possible assumptions regarding present-biased preferences: The first, that naive individuals do not acknowledge their self-control problem; the second, that sophisticated individuals fully understand their self-control problem. O'Donoghue and Rabin (2001) developed a model with partial naivete. A partially naïve individual estimates future present bias to be less severe than it will actually be. While we assume, for simplicity, sophisticated individuals and show ways that they could overcome self-control problems, our results could hold for partially naive individuals as well[3]. Note that in the context of our model, fully naive individuals would always have a self-control problem.

O'Donoghue and Rabin (1999) distinguish between two types of self-control problems, those displaying immediate rewards with delayed costs and those displaying immediate costs with delayed rewards. In this paper we focus on a task displaying immediate rewards with delayed costs.

[3] When a person estimates future present-bias to be sufficiently severe so that he understands that he will have a self-control problem in the future, the strategies discussed in the following sections to overcome self-control problems hold.

2 Self-Control Problems

Consider an individual who must choose, in each period t, whether or not to perform a task. There are an infinite number of periods in which the individual can perform the task. Each task $a_t(v,c) ? \{0,1\}$ is characterized by benefits, v, that are received in period t and costs, c, that are incurred in period t+1. When the individual does not perform the task in period t, $a_t(v,c) ? 0$; if he does perform the task, $a_t(v,c) ? 1$. We formalize the self-control problem by assuming that preferences are present-biased. Let u_t be individual instantaneous utility at time t. Let U^t, individual intertemporal preferences at period t, be represented by the following utility function[4]:

$$U^t(u_t, u_{t?1}, ...) ? u_t ? ? \sum_{?=t?1}^{?} ? ^{?=t} u_? \qquad \text{for all t,} \qquad (1)$$

where $0 ? ? ? 1, \quad ? ? 0$

The long run, time-consistent discounting is represented by $?$. When $? ? 1$, preferences are exponentially discounted. Present-bias is represented by $? ? 1$. We assume that in each period, the individual chooses whether to perform a task in order to maximize (1) while perceiving the number of periods as infinite. Present-biased utility function alters at each time period. Thus, we interpret the decision-maker to be a person having multiple selves and who evinces different incarnations in each time period. The individual will perform the task in each period iff $v ? ??c$. When $v ? ?c$, the individual will not want to perform the task in the future.

[4] Present-biased utility functions have been gaining popularity in recent years. Strotz (1956) and Pollak (1968) first used this form of utility function. Recent applications include Laibson (1997), O'Donogue and Rabin (1999,2001), Carrillo and Marriotti (2000) and others.

Definition: A self-control problem exists when $v - \beta\delta c$ and $v - \delta c$.

When an individual has a self-control problem he prefers to do the task in the present but not in the future. Since the decision whether to perform the task is made in each period, the individual does indeed perform the task in each period. Thus, his intertemporal utility is $v - \beta\delta c - \frac{\beta\delta(v - \delta c)}{1 - \beta\delta}$. When $v - \beta\delta c - \frac{\beta\delta(v - \delta c)}{1 - \beta\delta} < 0$, the individual will be willing, if he could, to forgo performing the task in the present in order not to perform the task in the future[5].

2.1 Pareto Superior Subgame Perfect Refinement for a Time Inconsistent Framework

In this section, we explore a difficulty with subgame perfect strategies in a time-inconsistent environment and propose a solution by using strategies that are Pareto superior. We define a history, h^t as a sequence of 0 and 1 from time period 0 to $t-1$ and define x as the sum of 1's in h^t, that is, the number of periods that the individual performed the task in the past. Consider an individual with an immediate-reward self-control problem ($v - \beta\delta c$ and $v - \delta c$) who previously performed the task during x periods. He contemplates whether to perform the task in the present. He might try never

[5] A self-control problem involving immediate costs is analogous to one involving immediate rewards. Assume an immediate-costs task with future rewards in the following period. The individual who maximizes (1) will perform the task iff $\beta\delta v > c$. However, he prefers to perform the task in the future iff $\delta v < c$. When $\beta\delta v > c$ and $\delta v < c$, the individual will never perform the task and has a self-control problem. His intertemporal utility is 0. If he would perform the task in every period, his intertemporal utility would be $\beta\delta v - c - \frac{\beta\delta(\delta v - c)}{1 - \beta\delta}$. When his intertemporal utility is positive he will be willing to pay up to his intertemporal utility for a commitment device that will commit him to perform the action in every period against his present-biased preferences.

to perform the task by following strategy D(x): In each period, do not perform the task if you performed the task during exactly x periods. Otherwise, perform the task.

PROPOSITION 1: *Using the D(x) strategy, it is a subgame perfect equilibrium not to perform the task when* $v - \beta c - \frac{\beta(v-c)}{1-\beta} \geq 0$ *and the individual performed the task during exactly x periods previously.*

Proof: See Appendix.

We view strategy D(x) as problematic. For any given x, it is optimal for an individual to perform the task and switch to a new strategy, D(x+1). A subgame perfect strategy lacks credibility in a time-inconsistent environment where the individual plays a game with his future incarnations because such a strategy does not capture the idea that, in each period, the strategies of all future players (i.e., the individual's future incarnations) or of a group of future players can be altered simultaneously. The reason strategy D(x) is not consistent is that it is not Pareto superior from the perspective of the future: In the future, the individual will wish to deviate to another, better, strategy. Peleg and Yaari (1973) studied changing preferences and define an equilibrium consumption path as follows:

> An equilibrium consumption path is optimal, in the sense that it has the following consistency property: the agent has no motivation to change his action in period t, nor does he have reason to regret his action in any period, given his actions in other periods. (P. 395)

We also look for an equilibrium from which no incarnation will wish to deviate[6]. We believe that an appropriate equilibrium for an individual who plays a game with his future incarnations can be a **Pareto superior subgame perfect equilibrium in each history of the game**[7]. Thus, in no time period is there a superior subgame perfect strategy. Such an equilibrium answers the dilemma we face in a time-inconsistent framework. On the one hand, the game is of an infinite number of players where each incarnation is represented by a different player. On the other hand, all players are linked; they can therefore coordinate a deviation to a superior strategy. When each incarnation receives a Pareto superior outcome from any possible subgame perfect equilibrium, no incarnation will wish to deviate. In order to offset a Pareto superior subgame perfect equilibrium we require only that there is some better subgame perfect equilibrium for any incarnation. We do not require the offsetting subgame perfect equilibrium to be Pareto superior.

In our simplified setting, all periods are the same. Any future utility function from its future point of view is exactly like the present utility function from the point of view of the present. Thus, if an action is best in the present perspective, it is also best in the future perspective.

The reason that there exists no Pareto superior subgame equilibrium is that the individual's first best option (perform the task in the present, do not perform the task in the future) is a subgame perfect equilibrium that is not a Pareto superior subgame perfect equilibrium. Any equilibrium is offset by the individual's first best option.

---

[6] Our agent always regrets his action because benefits are immediate, costs delayed and the utility function does not include past utility. Thus, we do not impose the regret condition.

[7] The notion is also close to the renegotiation proof literature. Yet, because a self-control context requires only one person, renegotiation can be seen as problematic. The literature on the renegotiation proof in infinitely repeated games includes: Farrell and Maskin (1989), van Damme (1989), Pearce (1988) and Abreu, Pearce, and Stachetti (1989).

PROPOSITION 2: *When the individual knows* h^t*, there exists no Pareto superior subgame perfect strategy.*

Proof: See Appendix.

2.2 Limited Memory

In order to allow viable Pareto superior subgame perfect strategies, we assume that the individual's cognition is limited in such a way that he does not remember the game's history.

Limited memory assumption: The individual does not remember whether he performed the task in any previous period.[8]

Motivation for the limited memory assumption is provided by Festinger's (1957) cognitive dissonance theory. Cognitive dissonance theory claims that when a person holds two conflicting cognitions, he experiences discord or tension and thus has an incentive to alter one of the cognitions. Gilad, Kaish and Loeb. (1987) describe a model where cognitive dissonance generates an information filter that blocks information when there is a discrepancy between new and old information. We believe that self-control situations are susceptible to such dissonance because an individual with a self-control problem experiences two conflicting cognitions. On the one hand, he believes that he should not perform the task over the long run. On the other, he prefers to perform the task at every period. If a person tells himself that he did not perform the task in the past, his dissonance is reduced.

[8] An alternative assumption can be that the individual knows whether he previously performed the task but, if he did perform it previously, does not remember when or how many times.

Under conditions of limited memory, the individual cannot devise strategies that will cause him not to perform the task because performing the task does not affect the known history in the future.

3 Strategies Used by a Wife to Sober Up Her Alcohol-Prone Spouse

In this section, we return to Fred and Wilma to describe how a relationship can help overcome an immediate-reward self-control problem. Fred has a drinking problem. Fred's utility function is specified by (1). During each period t, his utility from drinking is immediate and equal to v, while the costs of drinking are incurred in the next period and are equal to c. The self-control problem exists because $v ? ??c$ and $?c ? v$. If Fred could commit never to perform the task (i.e., never to drink), he would commit because $v ? ??c ? \frac{??(v ? ?c)}{1 ? ?} ? 0$. But Fred has no commitment device available. We assume that all Fred can do is talk to his time-consistent wife (Wilma) in all periods. Wilma cannot coerce Fred to do anything at any time; nor can she influence his preferences. Her objective is to convince Fred to act according to his long-term preferences. Wilma's time-consistent intertemporal utility, derived from Fred's instantaneous utility in each period, is specified by:

$$u_w^t(u_t, u_{t?1}, ...) ? ? \sum_{?=t}^{?} ? ^{??t} u_? \quad \text{where} \quad ? ? \{0,1\} \text{ and } j ? 0 \qquad (2)$$

Fred's instantaneous utility in time period $?$ is $u_?$. Wilma can be altruist toward Fred and care about his utility $(? ? 1)$ or not $(? ? 0)$. Wilma might also have a

preference for consistency. If she has such preference and goes back on her word in time period t, she incurs a psychological cost of j at time t.[9]

We assume that Wilma's objective is to maximize (2). Both Wilma and Fred perceive an infinite horizon future. Wilma is fully rational and remembers the game's entire history. Fred's cognition is restricted by the limited memory assumption specified in the previous section; he does not remember whether he drank in any previous period. An alternative to the utility function (2) stipulates that Wilma enjoys consumption from a fixed endowment available to her and Fred. The more Fred drinks, the less is left for other expenditures, which she values more than Fred's drinking. She thus does not want Fred to drink. Another alternative is that Wilma cares about Fred's actions rather then his utility. For our arguments to hold, we only require that Wilma want Fred not to drink and that she displays a preference for consistency.

3.1 A Silent-Treatment Strategy

Consider Wilma's *silent-treatment strategy* for sobering Fred up: In each period, tell Fred not to drink. If he does drink, do not talk to him about drinking for k periods[10]. Fred's *listening* strategy is: Listen to Wilma; if she tells you not to drink, do not. Else drink.

First we show that Fred's strategy is Pareto superior subgame perfect given Wilma's strategy. Then we study under what situations Wilma's strategy is credible.

[9] If Fred had a strong enough preference for consistency, he would not have a self-control problem. But we show that Wilma can convince Fred not to drink when j=0.

[10] Using this strategy, Wilma's function in the relationship is signaling. Wilma can thus persuade Fred not to drink by a simple action. In this section we assume common knowledge about the preferences that Fred and Wilma hold. In the next section we will relax this assumption.

LEMMA 1: *When the silent treatment lasts for a sufficient number of periods and Fred's discount rate between future periods is high enough, then Fred's listening strategy is a Pareto superior subgame perfect strategy and in equilibrium, Fred does not drink.*

Proof: See Appendix.

Fred does not suffer any disutility from the silent treatment per se. His utility function includes only utility from drinking. Rather, he knows that if Wilma does not talk to him for k periods, he will not have any mechanism to stop him from drinking during those periods. The number of periods required for Wilma not to talk to Fred if he drinks increases the more present-biased the preferences, the larger the rewards and the smaller the task's costs.[11]

When Wilma cares about Fred's utility and her inconsistency costs are sufficiently large $\left(? \; ? 1 \text{ and } \; j \; ? \; \frac{(?c ? v)(? \; ?? \; ^{k ?1})}{1 ? ?} \right)$, she can persuade Fred not to drink by using the silent treatment strategy because the strategy is credible. If Fred drinks despite Wilma's request and threat, she has an incentive to forgive him but if she will do so, she will incur the inconsistency cost of j, which is higher than the disutility she incurs if he drinks.

[11] Note that even for k infinitely high we require that $\frac{?}{1 ? ?} \; ? \; \frac{v ? \; ??c}{? \, (?c \; ? \; v)}$. If there is a low enough discount rate, even an infinite number of silent periods will not convince Fred to prefer the present. When the discount rate between future periods is $? \;\; ? 1$, we still require $k \; ? \; \frac{v ? \; ?c}{? \, (c ? \; v)}$ for the strategy to sober Fred up.

Another situation where Wilma can persuade Fred not to drink is when she does not care about his drinking ($? ? 0$) because under such circumstances she has no incentive to forgive Fred if he does drink. A question then arises of why Wilma should go to the trouble of helping Fred if she does not care. We assume that people involved in a relationship are willing to help one another even when they are indifferent to the actual behavior. It is not necessary for Wilma not to care about Fred at all; we only require that her utility not be influenced by whether he drinks or not. But if she does not care, and Fred would like her to forgive him, why should she not? We incorporate the inconsistency cost that Wilma incurs if she goes back on her word. The inconsistency cost can be arbitrarily small to strengthen the equilibrium. Even without such inconsistency cost, Wilma's threat is credible in equilibrium. The inconsistency costs give Wilma a reason not to forgive Fred.

PROPOSITION 3: *Wilma can persuade Fred not to drink by implementing the silent treatment strategy for a sufficient number of periods when (1)? and k are sufficiently large, and when* $? ? 1$ *and* $j ? \frac{(?c?v)(? ??^{k?1})}{1??}$ *— that is, i.e., she cares about his drinking and has a sufficiently high inconsistency cost or (2)* $?$ *and k are large sufficiently large and* $? ? 0$ *— that is, the utility she derives from his utility is zero.*

Proof: See Appendix.

However, when Wilma cares about Fred's utility ($? ? 1$) and $j ? \frac{(?c?v)(? ??^{k?1})}{1??}$, she cannot persuade Fred to stop drinking by using the silent treatment strategy because the strategy is not credible. If Fred drinks despite Wilma's request and threat of not talking to him, she has an incentive not to implement the threat. Since Fred knows this, he

will drink. When Fred has a self-control problem and Wilma cares about his long-run utility, she might have a self-control problem in trying to stop Fred from drinking.

Assuming there are low inconsistency costs ($j ? \frac{(?c?v)(? ?? ^{k?1})}{1??}$), when Wilma is an uninterested party ($? ? 0$) she can persuade Fred not to drink. Thus, in order to overcome self-control problems, it may be preferable to have an uninterested principle. Institutions such as Alcoholics Anonymous, drug clinics, and bureaucratic committees might function as such uninterested principles.

3.2 The Value of not Knowing Your Partner's Preferences: Strategic Pretence[12]

In this section we will show how Wilma can convince Fred not to drink when she cares about Fred's utility and has low inconsistency costs ($? ? 1$ and $j ? \frac{(?c?v)(? ?? ^{k?1})}{1??}$) and Fred believes that there is a probability, however small, that Wilma can credibly persuade him not to drink[13] (either $j ? \frac{(?c?v)(? ?? ^{k?1})}{1??}$ or $? ? 0$). Such ambiguity allows Wilma to make a credible threat. Wilma knows of Fred's uncertainty. Fred's subjective evaluation of $?$ and j is such that they take the following distribution:

$$\begin{array}{llll} ? ? 0 & or & (? ? 1 \quad and \quad j ? \frac{(?c?v)(? ?? ^{k?1})}{1??}) & with \quad probability \quad p \\ ? ? 1 & and & j ? \frac{(?c?v)(? ?? ^{k?1})}{1??} & with \quad probability \quad 1 ? p \end{array} \qquad (3)$$

were $0 ? p ? 1$

[12] By pretence we mean pretending to have emotions that one does not actually have. These emotions do not necessarily need to be virtuous. Wilma cares but acts as if she does not.

[13] We assume common knowledge of that cognition. It is also enough for Wilma to think that Fred is unsure about her preferences, with common knowledge regarding that cognition. Or that Fred thinks that Wilma thinks that Fred is unsure, with common knowledge regarding that cognition. And so forth.

In this section we also assume that if Fred drinks, he *knows* that Wilma is not supposed to talk to him for k periods. In equilibrium, that knowledge turns out to be true[14].

If Wilma reveals that she does care and has low inconsistency costs, Fred will deduce that she cannot help him overcome his drinking problem because her threat is not credible; hence, he will drink in all time periods, a situation that is undesirable for Wilma. She will thus always act as if she does not care or cares and has high inconsistency costs and keeps her threat of not talking to Fred for k periods. Because Fred knows that Wilma's threat is credible, he will not drink.

PROPOSITION 4: *The silent treatment strategy for Wilma and the listening strategy for Fred are Pareto superior subgame perfect equilibrium when 0<p<1 —i.e., when Fred is uncertain whether Wilma cares about his utility from drinking and/or whether her inconsistency costs are small enough and ? and k are large sufficiently large.*

Proof: See Appendix.

When Fred is unsure about Wilma's preferences, Wilma might act differently than when Fred knows her preferences. Wilma exploits Fred's uncertainty to achieve a better payoff for both Fred and herself. The result also shows the benefit to Fred of doubting whether Wilma cares about his drinking problem and/or whether she has high or low inconsistency costs. Thus, receiving information regarding Wilma's true preferences might be of negative value for Fred.

[14] This knowledge is available to Fred during the next k periods. He can deduce from this knowledge that he drank during one specific period. If he drinks during the k periods, he will not remember when he drank because it does not induce a new punishment phase (i.e., Wilma's refusal to talk to him). We thus assume that Fred understands and remembers punishment phases.

PROPOSITION 5: *Fred does not want to receive information about how much Wilma cares about his drinking (?) and/or her inconsistency cost (j) when he is uncertain about those parameters (0<p<1), when Wilma implements the silent-treatment strategy for a sufficient number of periods and when ? is sufficiently large.*

Proof: See Appendix.

Proposition 5 is reminiscent of Carrillo and Mariotti's (2000) "strategic ignorance" result for hyperbolic discounting agents. Here, Fred prefers to stay ignorant so that Wilma's threat will be credible. The difference between the two cases lies in the fact that Fred already has a self-control problem whereas in Carrillo and Mariotti (2000), the agent prefers strategic ignorance in order to avoid a self-control problem.

3.3 Strategic Self-Punishment

Close relationships are situations where it might be natural to assume that each partner knows the other's preferences; thus, the remedy offered in section 3.2 might not hold. In this section we address situations where according to section 3.1 Wilma could not convince Fred not to drink; We assume that ? ? 1 and $j ? \frac{(?c?v)(???^{k?1})}{1??}$ — that is, Wilma does care about Fred's drinking, has low inconsistency costs and there is common knowledge regarding Wilma's preferences. We add another tool: Fred can inflict negative utility (σ self-punishment) upon himself in every period. Consider the following *silence punishment strategy* as an option for Wilma:

Tell Fred not to drink if Fred never drank before, or drank before but punished himself in the previous period, or did not drink since he last punished himself. Otherwise, tell him to punish himself r^{t-z-1} (Fred drank in period z, the current period is t, $\beta > \frac{1}{\gamma\gamma}$, and r is sufficiently large) and do not tell him not to drink.

Fred's *self-punish* strategy is: Don't drink when Wilma tells you not to. If you did drink, punish yourself r^{t-z-1} and do not drink in the present period.

Wilma must be indifferent as to whether Fred punishes himself. We also assume that the sequence of events within each period is: first, Wilma either tells Fred not to drink or tells him to punish himself; then, Fred either punishes himself or not; finally, Fred decides whether or not to drink. The self-punishment is desirable and employed ex-post, after the harm from drinking is incurred, so that future selves will not drink.

PROPOSITION 6: *When Wilma cares about Fred's utility from drinking and has low inconsistency costs but does not care if he punishes himself, she can credibly prevent him from drinking by using the silence punishment strategy when the initial punishment is sufficiently large,* $r > \max\left\{\frac{v-\gamma c}{\beta\gamma - 1}, \frac{v-\gamma c}{\gamma}\right\}$, *the punishment increase at an adequate rate,* $\beta > \frac{1}{\gamma\gamma}$ *and* γ *is sufficiently large. Thus, Wilma's silence punishment strategy and Fred's self-punish strategy are a Pareto superior subgame perfect equilibrium in which Fred does not drink.*

Proof: See Appendix.

Fred prefers not to drink than to drink and punish himself because the punishment is sufficiently large. If he does drink, he prefers to punish himself immediately because the punishment increases with time[15]. Wilma will not alter her strategy because she knows that if Fred drinks, he will punish himself and stop drinking. She thus does not have any incentive to forgive him. Wilma must be indifferent toward the punishment. If she receives positive utility from the fact that Fred punishes himself, she will demand that Fred punish himself even when he does not drink. In such a case, since Fred will drink at each period because he knows that Wilma will blackmail him regardless of whether he drinks or not. He will also not punish himself because the only incentive he has for self-punishment is not drinking in future periods. If Wilma receives negative utility from the fact that Fred punishes himself, once he drinks she has an incentive to forgive him and alter her strategy; hence, she tells Fred not to drink. Because Fred knows that she will forgive him, he drinks immediately. Although Wilma's indifference towards the punishment Fred inflicts upon himself is essential for equilibrium, Wilma always strictly prefers her actions.

The sequence is important because if Fred punishes himself first and only then does Wilma tell him not to drink. In a situation where Fred does not punish himself, Wilma will have an incentive to forgive Fred for the fact that he did not punish himself and to tell him not to drink. An alternative to the increasing punishment scheme can be

---

[15] Note that if the punishment scheme is to be modified, then the punishment can be $r ? \frac{v ? ??c}{??}$ by increasing $?$ sufficiently, rather than $r ? \max \left\{ \frac{?v ? ??c}{???? ? 1}, \frac{v ? ??c}{?? ?} \right\}$, as stated in proposition 6. A tradeoff can be achieved between the rate of increase and the absolute level of the punishment.

that Wilma tells Fred to punish himself for more periods, thus increasing the duration of the punishment rather than its intensity.

4 Concluding Remarks

Psychological research and introspection suggest that people have self-control problems. This essay explores ways in which individuals can overcome their self-control problems through utilizing relationships. We need not assume that the individual with a self-control problem evaluates the relationship per se. Rather, he only evaluates the relationship as far as it helps overcome his self-control problem. In the sense suggested here, one reason for relationships to be established might be to help overcome self-control problems. The model also shows that even people who behave as if they were time-consistent might be internally time-inconsistent. All actions (drinking or not, talking, silent spells and self-punishments) are enacted under discretion in the period they are preformed. As such, they arise from the fact that Fred and Wilma do not want Fred to enact a certain behavior — drinking — in the future.

When a person in a relationship does not act in an optimal manner, self punishments and silent spells can be used to return that person to optimal behavior. If we assume that when Fred tries not to drink, he will still drink with some positive probability then behaviors like silent spells and self-punishments can be described as equilibrium behaviors that occur with assignable probabilities.

Anger might also help Wilma persuade Fred not to drink when she cares about his drinking and has low inconsistency costs. Wilma's strategy is credible if during those

"angry" periods she does not care about Fred's utility or gets angry and the negative utility she derives from talking to him outweighs the negative utility derived from Fred's drinking. According to the Lemma 1, Fred will not drink when Wilma's strategy is credible. It is sufficient that Fred think that if he drinks, Wilma will be angry with a certain probability. In this case, the situation is as described in the strategic pretence section (3.2).

We do not model two possible strategic behaviors: first, that Wilma has an incentive to convince herself not to care about Fred or to have high inconsistency costs, and second, that Fred has an incentive to convince Wilma that he is unsure about whether she cares about him and whether her inconsistency costs are high. Endogenizing such strategic behaviors is left for further research.

The drinking problem is a self-control problem of immediate rewards and later costs. The strategies discussed can be applied to self-control problems involving immediate costs with later rewards. For instance, Wilma can persuade Fred to always perform a task by telling him to do so in each period. If he does not, she will not talk to him for some periods. The analysis is similar to the analysis for the immediate-rewards self-control problem.

It is evident that people also overcome self-control problems by themselves. An individual can overcome his self-control problem when his cognitions are such that he knows that he did not drink anytime in the past and that drinking once will limit his memory, thus he will not remember whether he performed the task in the past. If his memory is limited, his only subgame perfect equilibrium is to drink. In order not to limit his memory and drink during all future periods, he will not drink at present. If the

individual knows that he will forget how much he drank and when he drank if he drinks d times, then he will drink d-1 times and then quit. Such scenarios can be thought of as "abstentions as commitment devices" because the individual will abstain from drinking at present in order not to drink in the future. Another possibility is an intra-personal strategy similar to the one presented in Section 3, where a wife was sobering up her husband. Assume that there are two selves within a person. A superego self and a decision-making, present-biased self. The superego tells the decision-maker not to drink. If the decision-maker drinks, the superego is angry and does not talk to the decision-maker for some time. The decision-maker thus decides not to drink when he perceives the superego's threat as credible. Analyzing such strategic intra-personal interaction is also left to further research.

The paper deals with one person's self-control problem when another person is trying to help him overcome it. We conjecture that the framework can likewise be applied to a situation where two people have a mutual self-control problem, as when both partners want to consume more at each present period but not in the future. One person can be the decision-maker, the other a controller and together they can overcome their mutual self-control problem. Such a model might explain role-playing in a relationship, as when he drives and she tells him to drive safely or when she shops while he tells her not to overspend.

References

Abreu, D., Pearce, D. and E. Stachetti (1989) *Renegotiation and Symmetry in Repeated Games*. Mimeo, Harvard University.

Ainslie, G. (1991) Derivation of "Rational" Economic Behavior from Hyperbolic Discount Curves. *American Economic Review*, 81, 334-40.

Akerlof, G. (1991) Procrastination and Obedience. *American Economic Review*, 81, 1-19.

Battagline, M., Benabou, R., and J. Tirole (2001) *Self-Control in Peer Groups*. Mimeo. Princeton University.

Benabou, R. and Tirole, J. (2001) *Willpower and Personal Rules*. Mimeo. Princeton University.

Benabou, R. and Tirole, J. (2002) Self-Confidence and Personal Motivation, *Quarterly Journal of Economics*, 117(3), 871-915.

Benzion, U., Rapoport, A. and J. Yagil (1989) Discount Rates Inferred from Decisions: An Experimental Study. *Management Science*, 35, 270-84.

Carrillo, J and T. Mariotti (2000) Strategic Ignorance as a Self-Disciplining Device. *Review of Economic Studies*, 67, 529-544.

Elster, J. (2000) *Ulysses Unbound*. Cambridge, UK: Cambridge University Press.

Farrell, J. and E. Maskin (1989) Renegotiation in Repeated Games. *Games and Economic Behavior*, 1, 327-360.

Festinger, L. (1957) *A Theory of Cognitive Dissonance.* Stanford: Stanford University Press.

Frenderick, S. Loewenstein G. and T. O'Donoghue (2002) Time Discounting and Time Preferences: A Critical Review. *Journal of Economic Literature*, 40, 351-401.

Gilad, B., Kaish, S. and D. Loeb (1987). Cognitive Dissonance and Utility Maximization. *Journal of Economic Behavior and Organization*, 8, 61-73.

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. and R. J. Herrnstein (1995) Preference Reversals due to Myopic Discounting of Delayed Rewards. *Psychological Science*, 6, 83-89.

Laibson, D. (1994) *Essays in Hyperbolic Discounting*. Unpublished Doctoral Dissertation, Boston: MIT.

Laibson, D. (1997) Golden Eggs and Hyperbolic Discounting. *Quarterly Journal of Economics*, 112(2), 443-77.

O'Donoghue, T. and M. Rabin (1999) Doing It Now or Later. *American Economic Review* 89, 103-24.

O'Donoghue, T. and M. Rabin (2001) Choice and Procrastination. *Quarterly Journal of Economics* 116, 121-60.

Pearce, D. (1988) *Renegotiation-Proof Equilibria: Collective Rationality and Intertemporal Cooperation*. Mimeo. New Haven: Yale University.

Peleg, B. and M. Yaari (1973) On The Existence of a Consistent Course of Action When Tastes are Changing. *Review of Economic Studies*, 40, 391-401.

Phelps, E. S. and R. A. Pollak (1968) On Second-Best National Saving and Game-Equilibrium Growth. *Review of Economic Studies*, 35, 185-199.

Pollak, R.A. (1968) Consistent Planning. *Review of Economic Studies*, 35, 201-208.

Schelling, T. C. (1992) Self-Control. In G. Lowenstein and J. Elster (Eds.), *Choice Over Time*. New York: Russell Sage, 167-76.

Strotz, R. H. (1956) Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies*, 23, 165-180.

Thaler, R. H. (1981) Some Empirical Evidence on Dynamic Inconsistency. *Economic Letters*, 8, 201-207.

Thaler, R. H. (1992) *The Winner's Curse*. Princeton: Princeton University Press.

van Damme, E. (1989) Renegotiation-Proof Equilibria in Repeated Prisoner's Dilemma. *Journal of Economic Theory*, 47, 206-207.

Appendix

PROOF OF PROPOSITION 1: Assume that an individual is contemplating whether to perform the task in some period and he has performed the task in more than x periods. According to the strategy, he should perform the task during all future periods regardless of his current actions. Performing the task at present increases utility. He will thus choose to perform the task whenever he has performed it during more than x periods. Now assume that the individual is contemplating whether to perform the task after completing it only x times. He knows that if he performs the task at least once, it will be optimal for him to continue performing the task during all future periods; in this case, his utility will be $v ? ??c ? \frac{??(v ? ?c)}{1 ? ?} ? 0$. Yet, if he does not perform the task at present, according to his strategy, he will never perform the task and his utility will be zero. He will thus choose never to perform the task at any period. QED

PROOF OF PROPOSITION 2: Assume there is a Pareto superior subgame perfect strategy in which an individual does not perform the task in some specific period although he did perform it previously x times. When the period he is supposed not to perform the task arrives, he contemplates whether to do so nonetheless. He is better off performing the task at present and following the subgame perfect strategy D(x+1) in the next period, thus achieving his first-best payoff. Thus, the strategy is not Pareto superior subgame perfect. Next, assume that there is a strategy in which the individual always performs the task. He is better off performing the task only in the present period and following D(x+1) in the future. Because there is no Pareto superior subgame perfect strategy associated with the individual not performing the task at any period and no

Pareto superior subgame perfect strategy associated with the individual always perfoming the task, there exists no Pareto superior subgame perfect strategy whatsoever. QED

PROOF OF LEMMA 1: Assume that Fred drank and Wilma is not talking to him for k periods. Drinking during those periods increases instantaneous utility by $v ? ??c$ and cannot effect future actions because Wilma will not talk to Fred regardless and Fred will not remember whether he drank or not due to the limited memory assumption. The known history at future dates will therefore does not depend on current actions. It is optimal for Fred to drink at present. Now assume that Fred did not drink and Wilma is telling him not to drink in the current period. Fred's instantaneous utility from drinking would then be $v ? ??c$. The future costs from drinking at present are equal to the costs of drinking for k more periods, which are $?(?c ? v)\left(\frac{? ? ?^{k?1}}{1??}\right)$. Fred will thus choose not to drink when $\frac{? ? ?^{k?1}}{1??} ? \frac{v ? ??c}{?(?c ? v)}$, which holds for $?, k$ sufficiently large because the right-hand side of the inequality decreases with $?$ and the left-hand side increases (to infinity) with $?$ when k is sufficiently large. If Fred does not drink at the present period, according to his strategy, he will not drink at future periods. QED

PROOF OF PROPOSITION 3: When $? ? 1$ and $j ? \frac{(?c?v)(? ?? ^{k?1})}{1??}$, Wilma has an incentive to implement the strategy because, if she does not, Fred will drink at each period and she will receive negative utility of $\frac{?(v ? ?c)}{1 ? ?}$. When Fred does not drink, Wilma's utility is zero. Her strategy is credible because if she goes back on her word, she suffers disutility of j, which is higher than $\frac{(?c?v)(? ?? ^{k?1})}{1??}$, the disutility she suffers from Fred drinking for k periods. When $? ? 0$, Wilma is indifferent to implementing the strategy. We assume that she does implement it. Wilma's strategy is credible because she does not care about Fred's utility and thus receives utility of zero if she sticks to her strategy and utility of –j if she goes back on her word. Thus there is no superior strategy for Wilma at any period. According to the Lemma 1, Fred will not drink when Wilma's strategy is credible. QED

PROOF OF PROPOSITION 4: Assume Wilma does not care about Fred's drinking or cares and has high inconsistency cost. If Fred does drink, Wilma will not talk to him for k periods regardless of whether he drinks during those periods, according to proposition 3. Now assume that Wilma does care about Fred and has low inconsistency costs. If Fred does drink, she has an incentive to forgive him and talk to him immediately. However, she knows that if she does so, Fred will deduce that she cares and has low inconsistency costs; he will therefore drink forever. In this case, her utility will be $\frac{?(v ? ?c)}{1 ? ?}$. If she acts as if she does not care about him or cares with high inconsistency costs and does not talk

to him for k periods, her utility will be $\frac{(v ? ?c)(? \; ? \; ?^{k?1})}{1 ? ?}$, which is greater than $\frac{?(v ? ?c)}{1 ? ?}$. Therefore, Fred knows that if he drinks, Wilma will not talk to him for k periods with probability 1. Fred's strategy is Pareto superior subgame perfect when Wilma's strategy is credible according to Lemma 1. Wilma has an incentive to implement the strategy because in equilibrium, Fred does not drink and she receives her first-best payoff. QED

PROOF OF PROPOSITION 5: Fred does not want to drink in the long run and does not drink when he is uncertain about Wilma's preferences according to proposition 4. His intertemporal utility is therefore zero. If he acquires information that Wilma does not care about him ($? \; ? \; 0$) or has high inconsistency costs ($j ? \frac{(?c ? v)(? \; ?? \; ^{k?1})}{1?? }$), he will gain nothing and continue not to drink. If he acquires information that Wilma does care about him ($? \; ? \; 1$) and has low inconsistency costs $j ? \frac{(?c ? v)(? \; ?? \; ^{k?1})}{1?? }$, he will start to drink and his utility will be $v ? ??c ? \frac{?? (v ? ?c)}{1 ? ?} ? 0$. He will thus prefer not to acquire information about Wilma's preferences. QED

PROOF OF PROPOSITION 6:

We must show that the strategy dictates the best action for Fred and Wilma from the perspective of each period. We will prove the proposition in 4 stages: 1) Fred would prefer not to drink than to drink and punish himself. This occurs when the utility from

drinking at present and self-punishment, r, at the next period are smaller than zero, which is the utility if Fred does not drink. The inequality is: $v ? ??c ? ??r ? 0$; thus, $r ? \frac{v ? ??c}{??}$. If the inequality holds, then it certainly holds for a punishment in some future period. Assume that Fred decides to punish himself in some future period n. Until that period, he will drink because drinking increases each present utility and he will punish himself in the future regardless. The current period is t. His utility will thus be $v ? ??c ? ?(v ? ?c)\frac{? ? ?^{n?t}}{1 ? ?} ? ??? ^{n?t?1} r ? 0$. If the inequality holds for n=q, then it certainly holds for n>q. The reason is that the punishment increases in time and until the punishment, the individual drinks, which adds to the negative utility from deciding to drink at present and punishing oneself later.

2) If Fred drinks, he will prefer to punish himself immediately. That will occur when a) He prefers to punish himself immediately rather than never punishing himself and drinking forever. Thus, $? r ? v ? ??c ? ?(v ? ?c)\frac{?}{1 ? ?}$, which holds for $?$ sufficiently large; and b) He will prefer to punish himself immediately rather than in any future period t+h; h>1 when: $? r ? v ? ??c ? ??^{h} ? ^{h} r ? ?(v ? ?c)\frac{? ? ?^{h}}{1 ? ?}$. The right-hand side decreases with h, so Fred will always prefer to punish himself earlier than later. For h=1, $r ? \frac{v ? ??c}{??? ? 1}$ and, because we know from (1) above that r must be positive, we require that $? ? \frac{1}{??}$, which means that the punishment must increase at a greater rate than the individual's discount rate between the present and following period.

3) Wilma prefers that Fred punish himself rather than altering her strategy. Wilma is indifferent toward the punishment. Assume that Fred drinks. She knows that Fred will punish himself immediately and not drink in any future period. If she does not alter her strategy, she receives a utility of zero, which is the highest utility she can achieve.

4) Wilma will implement the strategy because it provides her with the first-best payoff. QED