DISCUSSION PAPER SERIES

No. 2680

PROMISES, PROMISES…

Juan D Carrillo and Mathias Dewatripont

INDUSTRIAL ORGANIZATION
AND PUBLIC POLICY

Centre for Economic Policy Research

www.cepr.org

An online version of this Paper can be found at www.cepr.org/pubs/dps/DP2680.asp

ISSN 0265-8003

PROMISES, PROMISES…

Juan D Carrillo, ECARES, Université Libre de Bruxelles and CEPR
Mathias Dewatripont, ECARES, Université Libre de Bruxelles and CEPR

Discussion Paper No. 2680
January 2001

Centre for Economic Policy Research
90–98 Goswell Rd, London EC1V 7RR, UK
Tel: (44 20) 7878 2900, Fax: (44 20) 7878 2999
Email: cepr@cepr.org, Website: www.cepr.org

CEPR Discussion Paper No. 2680

January 2001

ABSTRACT

Promises, Promises…*

This Paper considers a time-inconsistent individual who has the ability to make promises that lead to a financial or reputation loss if broken. We first identify conditions under which promises made are kept, and conditions under which they are (partially) broken. Second, we endogenize the financial loss from breaking promises by considering interpersonal monitoring and explicit contracting. We describe optimal contracting under the assumptions that monitoring requires meeting and that meeting also opens the door to renegotiation of earlier promises. Third, we show how the loss from breaking promises can be reinterpreted in terms of reputation loss in the presence of incomplete information. Finally, we argue that the above results remain valid when we replace time-inconsistent preferences with limits to contracting as the source the individual's commitment problem. This significantly enhances the generality of these results.

JEL Classification: A12, D84
Keywords: hyperbolic discounting, limits to contracts, promises, time inconsistency

Juan D Carrillo
ECARES, Université Libre de Bruxelles
CP 114
50 Avenue Franklin D Roosevelt
1050 Brussels
BELGIUM
Tel: (32 2) 650 4214
Fax: (32 2) 650 4475
Email: carrillo@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=131469

Mathias Dewatripont
ECARES, Université Libre de Bruxelles
CP 114
50 Avenue Franklin D Roosevelt
1050 Brussels
BELGIUM
Tel: (32 2) 650 4217/4
Fax: (32 2) 650 4475
Email: mdewat@ulb.ac.be

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=104808

*Part of this work was done while Mathias Dewatripont was visiting MIT. The authors are grateful to Isabelle Brocas and participants at the Harvard/MIT theory seminar, Toulouse, ULB and Lisbon for comments.

Submitted 23 November 2000

NON-TECHNICAL SUMMARY

The literature on time-inconsistent preferences analyses the behaviour of individuals who overemphasize instant gratification relative to distant pay-offs. It is shown that, under this type of preferences, individuals tend to underprovide effort in unpleasant tasks with delayed rewards. If we assume that individuals are fully aware of their time-inconsistent behaviour, *commitment* to future actions is the most elementary way to impose present desires. Naturally, commitment devices are not always available. A recent strand of this literature investigates different commitment technologies that help the individual in achieving his current goals (self-restraint strategies, investment in illiquid assets, investment in inefficient technologies, strategic ignorance, deadlines, etc.).

The present research focuses instead on *promises*, another quite natural commitment device against individual time inconsistency. It is indeed intuitively plausible to try and get around one's time-inconsistency by 'making promises' (to one's boss, family, friends etc.) not to overconsume, or to work hard. Such promises alleviate self-control problems if not fulfilling them results in some loss for the individual. In this context, our goal is threefold. First, we study the optimal use of promises. Second, we provide microeconomic foundations for the effectiveness of such contractual promises. Third, we show that the above results remain valid when we replace time-inconsistent preferences by limits to contracting as the source of the commitment problem of the individual. Our main results are summarized as follows.

First of all, we investigate when promises will be made and which form they will take. We are interested in the extent to which promises are kept and show that the answer to this question depends on the functional form of the 'cost of breaking the promise'. When the marginal cost is increasing in the 'size' of the failure to meet the promise, then an equilibrium effort slightly lower than the one promise is relative inexpensive, while big departures are the expensive ones. In this case, the individual makes promises they know in advance will not be kept, but that at least force them to increase the effort relative to the future desired level. By contrast, if the marginal cost is decreasing in the size of the failure to meet the promise, individuals only announce promises that will be kept. In a dynamic context, we also prove that the individual may find optimal at each date not to fulfil all the promises made in the past and yet engage in new promises for the future.

In a second step we provide microeconomic foundations for the effectiveness of promises. After all, in a rational expectation world, any deviation from the promise should be perfectly anticipated by every individual. This could make the promise ineffective in the first place.

First, we show that it is possible to get around this argument by considering explicit contractual promises. The commitment value of interpersonal relations comes from the fact that *interpersonal renegotiation is more difficult than intrapersonal renegotiation*; the former requires coordination between two parties, while the latter does not. We consider a situation where a time-inconsistent producer has to exert effort in a future date and asks a time-inconsistent monitor to check this effort. Agents agree on a transfer payment from the producer to the monitor whenever the former has been caught shirking. In this setting the optimal frequency of meetings is the result of the following two effects: first, lowering this frequency is costly because it reduces the opportunities of monitoring. Deciding never to meet provides full commitment against renegotiation but also destroys all the discipline provided by monitoring. By contrast, however, meeting all the time also destroys monitoring, this time through renegotiation. Hence, the optimal frequency of meetings is well defined and it allows individuals to partially (but not totally) get around their time inconsistency.

Second, we offer another microfoundation for the effectiveness of promises. This one is based on incomplete information about the cost of exerting effort. Think of the time-inconsistent individual as a seller who, through his effort level, chooses the 'quality' of an input that has to be later used by a buyer. The seller can make a promise early on concerning this quality, but this promise is cheap talk, and the buyer knows it. When it is time for the seller to exert effort, the buyer simultaneously has to choose a technology that will be the best 'fit' for the quality of the input that will be provided by the seller. We assume that, *ceteris paribus*, the seller incurs an *ex post* loss whenever the buyer makes a technology choice that 'counts' on a higher quality input than what the seller has decided to provide. In the presence of private information about the seller's effort cost, the buyer does not know *a priori* which technology is appropriate. As a result, a pooling equilibrium may exist, whereby high-cost sellers can benefit from pooling with low-cost ones in their promises. Indeed, this can induce the buyer to make a more 'ambitious' technology choice. As before, the promise serves as a commitment device for the high-cost sellers to exert more effort and produce higher quality, even if it may carry the cost of falling short of the promised quality.

In the last part of the Paper we argue that the source of the potential need for promises is the commitment problem for the individual due to their time-inconsistent preferences. However, as is well known in economics, commitment problems can arise even with standard time-consistent preferences: in strategic situations, tying one's hands in advance can be helpful as a way to influence the behaviour of others. This is for example the case whenever an individual has to contract in the presence of information problems and has to bear a positive share of the resulting inefficiency loss. It is then easy to show how our earlier results can be reinterpreted in terms of limited contracting, and thus, whether we take the explicit monitoring or the reputation-based foundations for the effectiveness of promises.

1 Introduction

In a seminal paper on time-inconsistent preferences, Strotz (1956) analyzes the behavior of an individual who overemphasizes instant gratification relative to distant payoffs. Under this type of preferences, the individual tends to underprovide effort in unpleasant tasks with delayed rewards. The problem is especially appealing given that experiments conducted both by psychologists (Ainslie (1975) and Mazur (1987) among others) and economists (Thaler (1981) and Bleichrodt and Johannesson (2000) among others) suggest that animals and humans exhibit this "salience for the present".[1]

If we assume that individuals are fully aware of their time-inconsistent behavior, *commitment* to future actions is the most elementary way to impose present desires, as Ulysses did in his famous encounter with the Sirens. Naturally, commitment devices are not always available. A recent strand of the time-inconsistency literature investigates different commitment technologies that may help the individual in achieving his current goals. First, Caillaud, Cohen and Jullien (1996) study the strategy followed by a time-inconsistent individual who can "self-restrain" his future choices. The paper proposes a new equilibrium concept in which the set of deviations is restricted to those strategies in which the individual will not have a further incentive to deviate. In the unique equilibrium of this one-person game, the individual optimally succeeds at each period in moderating his consumption. Laibson (1997) shows that investments in illiquid assets can prevent individuals from incurring in inefficiently high levels of consumption. This provides a rationale for the existence of Christmas Clubs and other assets characterized by both high illiquidity and low rates of return. Jovanovic and Stolyarov (2000) argue that in order to avoid spending too many hours at work in their future labor life, time-inconsistent individuals may choose a low-paying occupation. This commitment device may then result in the rejection of technological progress by the entire society. There is also a series of papers (Carrillo and Mariotti (2000), Brocas and Carrillo (1999,2000a)) where self-commitment is achieved through strategic ignorance. For example, a researcher with meager but encouraging information on the prospects of a difficult project may optimally stay away from further costless information and undertake it. Indeed, extra knowledge may cast some doubts about the quality of the project and, because of his time-inconsistent preferences, lead the agent to beliefs involving inefficient procrastination. Last, O'Donoghue and Ra-

[1] See Ainslie (1992) and Loewenstein and Prelec (1992) for a comprehensive empirical and theoretical comparison of time-consistent preferences (exponential discounting) and time-inconsistent preferences (hyperbolic discounting) and Rubinstein (2000) for a criticism of the experimental evidence.

bin (1999) show that is optimal for a time-consistent principal to specify a deadline when contracting with a boundedly rational, time-inconsistent agent who is totally or partially unaware of his self-control problem.

The present research focuses instead on *promises,* another quite natural commitment device against individual time-inconsistency.[2] It is indeed intuitively plausible to try and get around one's time-inconsistency by "making promises" (to one's boss, family, friends, etc.) not to overconsume, or to work hard. Such promises alleviate self-control problems if not fulfilling them results in some loss for the individual. This loss can be random, through a probability of being caught shirking, or deterministic, if failing to meet the promise means finishing the job late. It can represent a financial loss, if there is a penalty for shirking, or a reputation loss, if a late job affects the individual's future reliability. In this context, our goal is both to study the optimal use of promises (Section 2) and to provide a microeconomic foundation for the effectiveness of such contractual promises (Section 3). Finally (Section 4), we argue that the above results remain valid when we replace time-inconsistent preferences by limits to contracting as the source of the commitment problem of the individual. Our main results are summarized as follows.

First of all, we investigate in Section 2 when promises will be made and which form they will take. We are interested in the extent to which promises are kept and show that the answer to this question depends on the functional form of the detection probability or reputation loss of shirking (from now on we will simply refer to this function as the cost of breaking a promise). When the marginal cost is increasing in the "size" of the failure to meet the promise, then an equilibrium effort slightly lower than the one promise is relative inexpensive, while big departures are the expensive ones. In this case, the individual makes promises he knows in advance he will not keep but that, at least, will force himself to increase the effort relative to his future desired level.[3] By contrast, if the marginal cost is decreasing in the size of the failure to meet the promise, individuals only announce promises that will be kept. Unfulfilled promises may also occur when the marginal cost is decreasing, as long as there is some uncertainty about the amount of effort specified by the promise. In a dynamic context, we also prove that the individual

[2]The approach to the issue of promises is very different from the one in Holmström and Kreps (1995). They focus on time-consistent individuals and assume away costs of breaking promises. In their setup, a promise is "cheap talk". It can be useful as a way to transmit information about a player's type. We also refer to Ellingsen and Johannesson (2000) for experimental evidence on the value of cheap-talk promises and threats.

[3]Note that we focus in this paper on individuals who, beyond their time-inconsistency, are fully rational or, to follow the O'Donoghue-Rabin (1999) terminology, "sophisticated".

may find optimal at each date not to fulfill all the promises made in the past and yet engage in new promises for the future. Finally, we briefly reinterpret the model in terms of inventory management, where an individual facing a known future demand has to decide first which inventory to buy as a contribution to serving demand and then how hard to work towards meeting the remaining unsatisfied demand.

In a second step, in Section 3, we provide microeconomic foundations for the effectiveness of promises. After all, in a rational expectation world, any deviation from the promise should be perfectly anticipated by every individual. This could make the promise ineffective in the first place.

In section 3.1, we show that it is possible to get around this argument by considering explicit contractual promises. The problem we consider is the following. A time-inconsistent agent (the producer) has to exert effort in a future date and asks another time-inconsistent agent (the monitor) to check this effort. Both agents agree on a transfer payment from the producer to the monitor whenever the former has been caught shirking. A good monitoring scheme is one where the contractual probability of being monitored is high and is also credible, i.e. renegotiation-proof. We assume that both monitoring and renegotiation require a physical meeting between the parties. We assume that, due to limited time availability, meetings have to be arranged in advance. This gives the individuals an ex-ante commitment power on the frequency of meetings. However, when they do get together, they cannot commit not to renegotiate the prescribed effort level. This assumption formalizes the idea that *interpersonal renegotiation is more difficult than intrapersonal renegotiation*: interpersonal renegotiation requires coordinating a meeting, which is time-consuming especially since it means freeing up time collectively.

In our model, the optimal frequency of meetings is the result of the following two effects: first, lowering this frequency is costly because it reduces the opportunities of monitoring. Deciding never to meet provides full commitment against renegotiation but also destroys all the discipline provided by monitoring. By contrast however, meeting all the time also destroys monitoring, this time through renegotiation. In our setup, the optimal frequency of meetings is well-defined and it allows individuals to partially get around their time-inconsistency. Interestingly, this approach highlights a case where interpersonal relations partially help solving individual incentive problems. While under time-consistent preferences interpersonal relations are at best neutral in terms of incentives and typically a source of problems, in our framework collective relations can improve

individual behavior.[4]

In Section 3.2, we offer another microfoundation for the effectiveness of promises. This one is based on incomplete information about the cost of exerting effort. Think of the time-inconsistent individual as a seller who, through his effort level, chooses the "quality" of an input that has to be later used by a buyer. The seller can early on make a promise concerning this quality, but this promise is cheap talk, and the buyer knows it. When it is time for the seller to exert effort, the buyer simultaneously has to choose a technology which will be the best "fit" for the quality of the input that will be provided by the seller. We assume that, ceteris paribus, the seller incurs an ex-post loss whenever the buyer makes a technology choice that "counts" on a higher quality input than what the seller has decided to provide.

In the absence of incomplete information about the seller's cost, the two individuals play a Nash equilibrium in effort and quality, and it can easily be shown that the earlier promise does not change the outcome. Time-inconsistency means that the seller would like to commit to higher effort than what he will in the end exert. However, the buyer will understand the seller's incentives to exert effort and will appropriately "scale down" the technology choice. Promises are thus ineffective. In the presence of private information about the seller's effort cost, the situation is different because the buyer does not know a priori which technology is appropriate. If the seller is time-inconsistent, a pooling equilibrium may exist whereby high-cost sellers can benefit from pooling with low-cost ones in their promises. Indeed, this can induce the buyer to make a more "ambitious" technology choice. Just as in Section 2, the promise then serves as a commitment device for the high-cost sellers to exert more effort and produce higher quality, even if it may carry the cost of falling short of the promised quality.

Section 3 thus provides two different microfoundations for the cost of failing to deliver on earlier promises, thereby validating the usefulness of promises as a commitment device against individual time-inconsistent preferences. In Section 4, we argue that the above results remain valid when we replace time-inconsistent preferences by limits to contracting as the source of the commitment problem of the individual. Indeed, what our results rely on is the fact that the individual starts with a commitment problem, which is the source of the potential need for promises. As is well-known in economics, commitment

[4]Naturally, we must be careful with what is considered an "improvement": a promise *increases* intertemporal welfare from the perspective of the individual who makes it, but it is *detrimental* for the future incarnation who is constrained to behave suboptimally from his viewpoint.

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 information problems (moral hazard or adverse selection) and has to bear a positive share of the resulting inefficiency loss. In Section 4, we show how our earlier results can be reinterpreted in terms of limited contracting, and this whether we take the explicit monitoring or the reputational foundations for the effectiveness of promises. In other words, in a world of limited contracting, promises can emerge as specific useful contractual instruments.

This 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. In a general perspective, it means that the literatures on time-inconsistent preferences and on limits to contracting could usefully learn more from one another. And, as far as this paper is concerned, this connection significantly enhances the generality of the idea that promises can be made even when one knows ex-ante that they will be broken partially at a cost.

2 Time-inconsistent preferences and promises

We analyze the behavior of an individual with time-inconsistent preferences, in the sense of Strotz (1956). This implies that current payoffs are *overweighed* (or *salient* in the words of Akerlof, 1991) relative to future payoffs. Using the standard notation introduced by Phelps and Pollak (1968), we posit that from the perspective of the individual at date t, period $t+s$ ($s \geq 1$) is discounted at a rate $\beta\delta^s$ where $\beta \leq 1$. Naturally, $\beta = 1$ is the standard case of exponential discounting, and therefore time-consistent preferences. Without loss of generality and except otherwise stated, we will assume that $\delta = 1$, and we will call self-t the incarnation of the agent at date t. Furthermore, we assume that the agent is "sophisticated" (that is, aware at every period of his self-control problem).

At date 1, the individual will be required to put some effort e in order to complete a given task. The cost of this effort is immediate and equal to $\psi(e)$, where $\psi'(e) > 0$ and $\psi''(e) > 0$. The benefit comes one period later, at date 2, and has a value e (output is deterministic and equal to effort). Given the individual's discounting, the surplus of self-0 and self-1 are respectively given by:

$$\beta\,[e - \psi(e)] \qquad \text{and} \qquad \beta\, e - \psi(e) \tag{1}$$

As one can easily see, the optimal effort that self-0 would like to exert at date 1 is strictly higher than the level of effort that self-1 is effectively willing to exert when date 1 arrives. Time-inconsistency (captured through the parameter β) is the source of an intrapersonal incentive problem.

Assume now that self-0 can make a "promise" e^* about the effort e at date 1, and that failure to reach this promised effort level e^* leads to a cost $f(e^*-e)$ for any effort e exerted at date 1. There are two alternative interpretations of this function $f(e^*-e)$. It may represent either the probability of being detected shirking, or the reputation loss from not being reliable. It is natural to assume that making a promise is costless as long as it is fulfilled or exceeded. When promises are not met, the cost is positive and increasing in the difference between the promise and the effort realized. This is summarized as follows.

Assumption 1 *$f(e^*-e)\equiv 0$ for all $e\geq e^*$ and $f'(e^*-e)>0$ for all $e<e^*$.*[5]

Note that $f(\cdot)$ convex (resp. concave) means that the marginal probability of detection or the marginal reputation loss is increasing (resp. decreasing) in the difference between promise and effort exerted. That is, small departures from the effort promised are relatively costless (resp. costly). For the time being we take the function $f(\cdot)$ as exogenously given. In Section 3 we provide foundations for this cost structure based either on probability of detection or on reputation loss vis-à-vis other agents.

The timing of the game can be summarized as follows.[6]

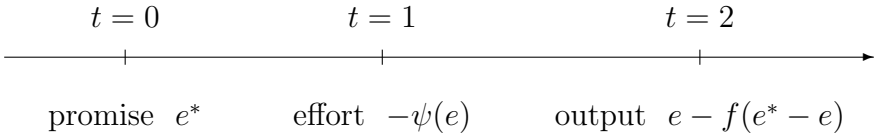


Figure 1. Timing.

2.1 Fulfilled vs. partially fulfilled promises

In our setting, self-0's intertemporal utility function is:

$$W(e,e^*)=\beta\Big[e-f(e^*-e)-\psi(e)\Big] \tag{2}$$

[5]We could also assume the function f to be increasing in $|e^*-e|$. This would leave most of our results unaffected.

[6]It does not matter when the cost of breaking a promise is paid. For simplicity, we assume that it occurs at the date in which output is produced.

But, because of the dynamic inconsistency of preferences, when the date of exerting effort arrives self-1's intertemporal utility function becomes:

$$V(e, e^*) = \beta\Big[e - f(e^* - e)\Big] - \psi(e) \tag{3}$$

As already stated, from (2) and (3) we can notice that absent a possibility of promises ($e^* = 0$), self-1 ends up exerting too little effort from self-0's viewpoint. Formally, denote $e_0(e^*) = \arg\max_e W(e, e^*)$ and $e_1(e^*) = \arg\max_e V(e, e^*)$. We get:

$$e_1(0) < e_0(0)$$

with $\psi'(e_1(0)) = \beta$ and $\psi'(e_0(0)) = 1$.

Assume the only instrument that individuals have at date 0 for "forcing themselves" to exert a high level of effort is the promise. Naturally, this promise may be ex-post costly (whenever it is not fulfilled). Our first concern is to provide a full characterization of the optimal promise e^* given the functional form $f(\cdot)$ for the cost of breaking it.

What we are considering is technically equivalent to a moral hazard problem, with self-0 setting the promise e^* as an incentive scheme for self-1. As is well-known, moral hazard problems are easily plagued by nonconcavity of the overall maximand.[7] In order to avoid that, we rely on the following technical assumptions:

Assumption 2 $\beta\, f''(e^* - e) > -\psi''(e) \quad \forall\, e^*$ *and* $e \leq e^*$.

Assumption 3 $f'''(e^* - e) \leq 0$ *and* $\psi'''(e) \geq 0$.

The first assumption says that $f(\cdot)$ is not "too concave" relative to $-\psi(\cdot)$. It guarantees that $V(e, e^*)$ is concave in the effort e exerted by self-1. The second one says that the rates of concavity of $f(\cdot)$ and $-\psi(\cdot)$ are nondecreasing in their arguments. It guarantees that $W(e_1(e^*), e^*)$ is concave in the promise e^* made by self-0 even when $f'' > 0$.

Denote by $\hat{e}$ the optimal level of effort exerted by self-1 *conditional* on the promise e^* being fulfilled (i.e. given $e^* = \hat{e}$). Naturally, this effort will depend on the marginal cost of a departure from the full promise $f'(0)$. Formally,

$$\left.\frac{\partial V(e, \hat{e})}{\partial e}\right|_{e=\hat{e}} = 0 \quad \Leftrightarrow \quad \psi'(\hat{e}) = \beta[1 + f'(0)]$$

Using this definition, we are in a position to state our first result.

[7] See e.g. Grossman and Hart (1983).

Proposition 1 *Consider the game where self-0 makes a promise and self-1 exerts effort. If the intrapersonal conflict is sufficiently small or if a departure from the promise is sufficiently costly, then self-0's optimal effort level is exerted by self-1 and promises are fulfilled (case* (i)*). Otherwise, self-0's optimal effort level is not reached. In that case, promises will remain unfulfilled if the marginal cost is sufficiently increasing in the difference between promise and effort (case* (iii)*) and they will be fulfilled if it is not (case* (ii)*). Formally,*[8]

(i) *If* $f'(0) > (1-\beta)/\beta$ *(i.e.* $\hat{e} > e_0(0)$*), then* $e^* = e_0(0)$ *and* $e_1(e^*) = e^*$.

(ii) *If* $f'(0) < (1-\beta)/\beta$ *(i.e.* $\hat{e} < e_0(0)$*) and* $f''(0) < f'(0)\psi''(\hat{e})/\beta(1-\psi'(\hat{e}))$, *then* $e^* = \hat{e} < e_0(0)$ *and* $e_1(e^*) = e^*$.

(iii) *If* $f'(0) < (1-\beta)/\beta$ *(i.e.* $\hat{e} < e_0(0)$*) and* $f''(0) > f'(0)\psi''(\hat{e})/\beta(1-\psi'(\hat{e}))$, *then* $\hat{e} < e_1(e^*) < e_0(0)$ *and* $e_1(e^*) < e^*$.

Proof. Consider first the optimization at $t = 1$. The first possibility is the following: if $e^* \leq e_1(0)$, then by (3) the promise is not binding and we have $e_1(e^*) = e_1(0)$. Second, recall that the optimal effort from self-0's perspective is $e_0(0)$. Now, given a promise e^*, if there is an interior solution to (2), it must satisfy:[9]

$$V_1(e_1, e^*) = 0 \;\Rightarrow\; \beta[1 + f'(e^* - e_1)] = \psi'(e_1) \tag{4}$$

Given Assumption 2, $V_{11}(e_1, e^*) < 0$ so the SOC of our maximization problem is satisfied. From (4), $V_1(e^*, e^*) = \beta[1 + f'(0)] - \psi'(e^*) = \psi'(\hat{e}) - \psi'(e^*)$. So, the second possibility is the following: if $e^* \in (e_1(0), \hat{e})$, then $V(e, e^*)$ is increasing in e for all $e \in [0, e^*]$, and the promise will be fulfilled. Finally, taking the derivative of the first-order condition w.r.t. the promise yields:

$$V_{11}(e_1, e^*)\frac{\partial e_1}{\partial e^*} + V_{12}(e_1, e^*) = 0,$$

which implies:

$$\frac{\partial e_1}{\partial e^*} = \frac{\beta f''(e^* - e_1)}{\beta f''(e^* - e_1) + \psi''(e_1)}.$$

So, the third possibility is the following: when $e^* > \hat{e}$, then $\dfrac{\partial e_1}{\partial e^*} < 0$ if $f'' < 0$ and $\dfrac{\partial e_1}{\partial e^*} \in (0, 1)$ if $f'' > 0$.

[8] Note that a sufficient condition for the second inequality in case (ii) to hold is $f'' < 0$, and sufficient conditions for case (iii) to hold are $f'(0) = 0$ and $f''(0) > 0$.

[9] Subscript l in $V(\cdot)$ means partial derivative with respect to the l th argument.

We can now turn to the optimization at date $t = 0$. Obviously, the promise becomes binding only when $e^* \geq e_1(0)$. Then, self-0's optimization problem $\mathcal{P}$ amounts to:

$$\mathcal{P}: \quad \begin{cases} \max\limits_{e^*} \quad e^* - \psi(e^*) & \text{if} \quad e^* \in [e_1(0), \hat{e}] \\ \max\limits_{e^*} \quad e_1(e^*) - f(e^* - e_1(e^*)) - \psi(e_1(e^*)) & \text{if} \quad e^* > \hat{e} \\ \qquad \text{s.t.} \quad \dfrac{\partial e_1}{\partial e^*} = \dfrac{\beta f''(e^* - e_1)}{\beta f''(e^* - e_1) + \psi''(e_1)} \end{cases}$$

These two cases thus concern respectively the second and third possibilities above.

By definition of $\hat{e}$, we have

$$\hat{e} > e_0(0) \;\Leftrightarrow\; f'(0) > \frac{1-\beta}{\beta},$$

so we can now conclude. First, case (i) of the Proposition is obvious. For cases (ii) and (iii), the first-order condition of the maximization problem w.r.t. e^* yields:

$$\frac{\partial e_1}{\partial e^*}\Big(1 - \psi'(e_1)\Big) - f'(e^* - e_1)\left[1 - \frac{\partial e_1}{\partial e^*}\right] = 0.$$

Assumption 3 ensures that the SOC will be satisfied.[10] Note that, for $e^* = \hat{e}$, the promise will be fulfilled and the derivative of the maximand at this point w.r.t. e^* is:

$$\frac{\beta f''(0)}{\beta f''(0) + \psi''(\hat{e})}\Big(1 - \psi'(\hat{e})\Big) - f'(0)\frac{\psi''(\hat{e})}{\beta f''(0) + \psi''(\hat{e})}. \tag{5}$$

Assumption 2 ensures that the denominator is positive. By Assumption 3, this derivative will be decreasing in the promise from this point on. Then, if (5) is negative, we shall have a fulfilled promise $e^* = \hat{e}$ (case (ii)). Otherwise, we shall have a higher promise and a higher effort that however fails short of meeting the promise (case (iii)). □

The idea of the proposition is the following. A promise makes sense only if it requires an effort higher than self-1's preferred level $e_1(0)$, and it will never prescribe an effort higher than the optimal from self-0's viewpoint $e_0(0)$. If the marginal cost of any deviation from the prescribed effort is sufficiently high (i.e. if $f'(0)$ is high enough so that, for example, any amount of shirking is detected with high probability), then failing to meet the promise is too costly, and therefore self-0 only imposes targets that can be fulfilled. These targets may not necessarily imply that self-0's optimal effort level is reached. Indeed, when the

[10] The SOC is: $\frac{\partial^2 e_1}{\partial e^{*2}}\big(1 - \psi'(e_1) + f'(e^* - e_1)\big) - \psi''(e_1)(\frac{\partial e_1}{\partial e^*})^2 - (1 - \frac{\partial e_1}{\partial e^*})^2 f''(e^* - e_1)$. By Assumption 3, $\frac{\partial^2 e_1}{\partial e^{*2}} < 0$ which guarantees that the second-order condition is satisfied.

intrapersonal conflict is too important, an excessively demanding promise does not act as a commitment device for higher future effort. It is then more interesting to set mild promises that are fully honored. Now, suppose that failing to meet the target "by a little" is not too costly (i.e. f' not too high) but this marginal cost increases with the difference between effort promised and effort realized ($f'' > 0$). In this case, by setting higher and higher targets, the individual is committing to exert more and more effort, even though these promises are never fulfilled. Targets are then raised by self-0 until the (constant) gains of a higher commitment to effort are offset by the (increasing) costs of unfulfilled promises. In equilibrium, self-1 is detected with some probability or loses some of his reputation.

Let us now illustrate this result with two examples.

Example 1: Unfulfilled promises. Consider the quadratic case, where $f(e^* - e) = (e^* - e)^2/2$ and $\psi(e) = e^2/2$. In this case, $e_0(0) = 1$ and $e_1(0) = \beta$. Note that $f'(0) = 0$ and $f'' > 0$, so we are in case (iii) of Proposition 1. Maximizing the payoff at date 1 given a promise $e^* \geq \beta$ yields:

$$e_1(e^*) = \frac{\beta}{1+\beta}(1 + e^*).$$

Note that $e_1(e^*) < e^*$ for all $e^* > \beta$: whenever the promise exceeds self-1's optimal effort $e_1(0)$, the individual chooses not to fulfill the promise. However, raising the promise still raises future effort. Consequently, "excessive" promises are made. Specifically, the optimization at date 0 yields:

$$e^* = \frac{2\beta}{1+\beta^2} \quad \text{and} \quad e_1(e^*) = \frac{\beta + \beta^2}{1+\beta^2}$$

so that we have $e_1(0) < e_1(e^*) < e^* < e_0(0)$.

Example 2: Fulfilled promises. Consider now the linear case $f(e^* - e) = (e^* - e)$ and quadratic cost of effort $\psi(e) = e^2/2$. We have $f'(0) = 1$ and $f''(0) = 0$ so, according to Proposition 1, we are either in case (i) or case (ii). We then get:

If $\beta > 1/2$ (i.e. $f'(0) > (1-\beta)/\beta$), then $e_1(e^*) = e^* = e_0(0) = 1$.

If $\beta < 1/2$ (i.e. $f'(0) < (1-\beta)/\beta$), then $e_1(e^*) = e^* = \hat{e} = 2\beta < e_0(0) = 1$.

In other words, the individual always fulfills his promises. Indeed, effort is set to match the promise until a certain threshold, and then stays constant. Setting the promise beyond this threshold thus makes no sense: it raises the probability of detection or reputation loss while failing to increase effort. As seen from the equation above, if the intrapersonal

conflict is sufficiently weak (β greater than 1/2) first-best effort is achieved. By contrast, if the conflict is strong enough (β smaller than 1/2) only second-best effort is achieved.

Examples 1 are 2 are graphically represented in Figure 2.

[INSERT FIGURE 2 HERE]

We conclude this section by considering two extensions of our framework, as well as one application. The reader mainly interested in the foundations behind our cost of broken promises can immediately turn to sections 3 and 4, which are entirely devoted to this subject.

2.2 Extension: random output

One might think of the promise as the commitment of self-0 to reach a certain target, in terms of quantity of good produced for example. If such output is used as an input of production by another agent, meeting the target might be essential in order to efficiently coordinate the activities of all individuals. This would amount to a highly concave function for the cost of breaking promises. Still, there are circumstances in which the total amount of production might partly depend on factors outside the control of the individual. We formally introduce this possibility by assuming that there is some uncertainty on the output produced and that, by exerting effort e, the individual affects its (stochastic) level. More concretely, output x is a random variable drawn from a Beta(2,2) distribution with support $[e, e+1]$ that depends on effort. From standard statistical theory, we know that its cumulative distribution function $G(x \mid e)$ is given by:

$$G(x \mid e) = 3(x-e)^2 - 2(x-e)^3 \qquad \forall\ x \in [e, e+1],$$

so the density function is symmetric around $e + 1/2$ and has an inverse U-shape:

$$g(x \mid e) = 6(x-e)(e+1-x) \qquad \forall\ x \in [e, e+1].$$

We suppose that the agent sells all the output produced. However, if he announces a target (or promise) e^* and this output is not reached, then he incurs in a fixed cost (normalized to 1) which is independent of the difference between the target and the output obtained. This extreme form of concavity for the cost function is summarized as follows.

$$\begin{cases} f(e^* - x) = 1 & \text{if} \quad x < e^* \\ f(e^* - x) = 0 & \text{if} \quad x \geq e^* \end{cases}$$

Naturally, in the absence of an intrapersonal conflict, promises are useless. Note also that, even if output is stochastic, by exerting an effort e greater or equal to the target e^* the agent is sure of never being short of output ($x \in [e, e+1]$). For any announced target e^* by self-0, the intertemporal welfare of a risk-neutral agent from his self-1's perspective is:

$$\begin{aligned} V(e, e^*) &= \beta \int_e^{e+1} x - f(e^* - x)\, dG(x \mid e) - \psi(e) \\ &= \beta \left[e + \frac{1}{2} - G(e^* \mid e) \right] - \psi(e) \end{aligned}$$

As we can see from the above equations, any target announced by self-0 results in a cost function which is first convex and then concave:

$$G'' > 0 \quad \text{if} \quad e^* - e \in (0, 1/2) \qquad \text{and} \qquad G'' < 0 \quad \text{if} \quad e^* - e \in (1/2, 1).$$

In this uncertain world, small differences between effort and promises are not very costly, since the risk of ending up with insufficient production is relatively small. However, as the difference starts growing, the chances of not meeting the target increases, and promises become more and more costly. The cost finally stabilizes when the likelihood of a short supply of output is sufficiently high ($e^* - e \geq 1/2$). Naturally, the results of Proposition 1 cases (i) and (iii), and in particular the fact that it might be optimal to set unreasonably high targets, hold in this new framework. To sum up, this section shows that unfulfilled promises may arise even when targets are rigid, as long as there is some uncertainty.[11]

2.3 Extension: repeated effort

The above one-shot problem can be extended to a context in which effort is chosen repeatedly. Specifically, call e_t the effort level chosen in period t and e_t^* the promise made for this period. As before, the effort cost is $\psi(e)$, while the cost of an unfulfilled promise is $f(e_t^* - e_t)$. We shall interpret e_t^* as the "stock" of yet unfulfilled promises at the beginning of period t. These are the promises just made in the previous period plus the excess of earlier promises over the efforts made in previous periods. We thus assume that an unfulfilled promise has a constant cost per period of delay.

[11] The main difference with the case where the cost is convex everywhere is that the individual will be more cautious in setting targets. In particular, the promise will never exceed the value $\bar{e}$ given by $\bar{e} - e_1(\bar{e}) = 1/2$. Besides, promises would be even less important if effort above the promise were costly (f increasing in $|e^* - e|$).

Each self-$t = 1, 2, ..., N$ (with N finite or infinite) will both choose a level of effort e_t and a promise e^*_{t+1}. Setting $\delta = 1$ as before, the maximization problem of self-0 is:

$$\max_{\{e_t, e^*_t\}} \quad \sum_{t=1}^{N} \; e_t - \psi(e_t) - f(e^*_t - e_t)$$
$$\text{s.t.} \quad e^*_{t+1} \geq e^*_t - e_t, \quad e_t \geq 0, \quad e^*_t \geq 0 \qquad \forall\, t.$$

Naturally, self-0 understands that future selves will distort this maximand when choosing their current effort level. In particular, they will overweigh the cost of their current effort by a multiple $1/\beta > 1$. Let us assume here the existence of a stationary interior solution with $e^*_t > e_t > 0$.[12] In such a case, self $t-1$ sets his promise under the expectation that e_t will satisfy:

$$\beta\Big[1 + f'(e^*_t - e_t)\Big] = \psi'(e_t).$$

This will imply that promise e^*_t will satisfy:

$$(1 - \psi'(e_t))\frac{de_t}{de^*_t} = f'(e^*_t - e_t)(1 - \frac{de_t}{de^*_t}).$$

The above two optimality conditions mean that current effort is solely determined by current costs and benefits, while next period's promise is determined solely by next period's costs and benefits. This is because, in an interior solution, the total value for the next promise is determined independently of the extent to which current effort fails to match existing prior promises.

Consequently, the above two equations are equivalent to those of the one-shot problem. This means that, if we make the functional form assumptions of Example 1 (that is, $f(e^*_t - e_t) = (e^*_t - e_t)^2/2$ and $\psi(e_t) = e_t^2/2$), we obtain as solutions:

$$e_t = \frac{\beta}{1+\beta}(1 + e^*_t) \qquad \text{and} \qquad e^*_t = \frac{2\beta}{1+\beta^2}.$$

It is interesting to notice that even if the agent never fulfills his promises ($e_t < e^*_t$), he still finds it optimal to commit to a higher future level of effort ($e^*_{t+1} > e^*_t - e_t$).

2.4 Application: inventory management

In this subsection, we offer a digression and reinterpret the above model in terms of inventory management, by considering an agent who faces a known future demand. The

[12] A full characterization of this dynamic solution is an interesting avenue for research, but is beyond the scope of this paper.

individual has to choose: (i) to what extent his demand should be met, and (ii) whether to meet it through inventories or through production. We thus look at a simple model in which both the level of inventories and the gap between demand and supply are endogenous. These issues have attracted a fair amount of attention in manufacturing production, for example in the context of the comparison of production methods in the car industry. As has been well-documented (see e.g. Womack et al., 1990), in the 1970s and 1980s the car industry has seen the emergence of "lean production", pioneered by Japanese carmakers, that gradually displaced the "mass production" invented in the US during the 1930s. Among its attributes, lean production includes a much lower level of inventories (the famous "just-in-time" system) and a higher level of productivity (cars per hour worked) and quality (as measured for example by consumer complaints).

The analysis sketched here is based on many simplifications and is therefore only suggestive. Consider an individual with time-inconsistent preferences who faces in period 2 a known demand d, at an exogenous unit price of 1. If he only manages to supply an amount $s < d$, he will suffer at date 2 an immediate revenue loss, but also a cost $f(d-s)$ due to his being perceived as "unreliable". His overall payoff is therefore $s - f(d-s)$. Supply s is the sum of two components: production e which takes place at date 1 and involves a cost $\psi(e)$, and prior inventory purchase $\bar{s}$ that has to be decided at date 0. Assuming that inventory can be bought at the same unit price of 1, so it does not yield any net contribution to revenue, the payoff of self-0 is:

$$\beta\Big[e - f(d - (\bar{s} + e)) - \psi(e)\Big].$$

Instead, self-1's payoff is:

$$\beta\Big[e - f(d - (\bar{s} + e))\Big] - \psi(e).$$

This problem is formally equivalent to the one described in (2) and (3), with the promise e^* being formalized here in terms of $(d - \bar{s})$, the gap between demand and prior inventory purchase. As far as the interpretation is concerned, note however that the demand level d (which is exogenous to the model) can be naturally considered as the "promise". The inventory $\bar{s}$ is self-0's choice variable, and it can be thought of as a way to *alleviate* the promise. Under this alternative interpretation and using the results of the previous subsection, we can draw several conclusions. First, while time-consistent individuals will make sure to keep inventories at a level that is sufficient to meet demand, time-inconsistent individuals may, provided we have $f'' > 0$, rely on insufficient inventories as a (costly) self-commitment device to raise production. Second, for a given level

of time-inconsistency, a cut in inventories simultaneously raises production and the gap between demand and supply: those who fail to meet demand by the bigger margin are also those who work harder. Third, a rise in time-inconsistency (i.e. a lower β) lowers production and tends to raise inventories. Conversely, a rise in β, i.e. a better alignment of the preferences of self-1 with the preferences of self-0 leads simultaneously to a lower level of inventories and higher production, which are the attributes of just-in-time manufacturing. In this respect, note that Section 4 will argue that a rise in β can be reinterpreted in terms of reduced agency problems in a world without time-inconsistent preferences but with incomplete contracting. Our model is thus quite consistent with the perceived differences between lean production and mass production, since it has been argued that lean production improves the relation between manufacturers and their suppliers.

3 Foundations for promises based on time-inconsistent preferences

In the previous section we have assumed the existence of an exogenous cost whenever a strategy announced by self-0 is "renegotiated" by self-1. We have then studied the optimal use of this tool in shaping future behavior. In particular, we have shown that commitment through promises is a way to avoid future procrastination and a level of production inefficiently low from the perspective of the current self. At the same time, it is a costly mechanism: excessively high targets which are ex-post not respected, may (optimally) be announced in equilibrium (see Proposition 1).

There is however an apparent tension between agents' perfect foresight and the existence of a cost of breaking promises. After all, if individuals in the economy are rational, a deviation by self-1 of the strategy announced by self-0 will be perfectly anticipated by all the parties. This will make the promise non-credible with respect to the outside world. How could it then be a useful tool for influencing future conduct? To be concrete, if for example insufficient supply of output relative to the level agreed is perfectly anticipated, how can there be an ex-post loss by the individual who engages in such behavior?

In this section we present two extensions of the basic model which show that this reasoning can be circumvented. Sticking to the time-inconsistency paradigm, we open the black box of the "cost of breaking promises" in two ways. In both cases, this leads to situations where time-inconsistent individuals make promises anticipating that they will break them with probability one in equilibrium. We show first that an individual can,

through interpersonal relations, ex-ante create an *endogenous financial cost* associated to the reneging on promises which is credible and influences ex-post behavior. Then, we move to an incomplete information setting where the individual makes promises (that he may later break) in order to entertain a reputation of high effort. Once again, this creates a credible *endogenous reputation cost* associated to the reneging on promises which usefully influences ex-post behavior. The reputation cost here comes from the fact that the outside world is not sure that the promise will be broken in equilibrium, because the individual that plans to renege on his promise pools with another individual that will honor it. Consequently, when a promise is broken ex-post, this was not fully anticipated ex-ante.

3.1 Financial cost of breaking a promise

3.1.1 Intrapersonal conflict and interpersonal contract

Consider the following extension of the model presented in Section 2. A time-inconsistent individual (the producer P) will be required to manufacture a good once between dates 1 and n. Production requires an immediate cost $\psi(e)$ and provides a one-period delayed benefit e. Given our previous formulation of time-inconsistent preferences ($\beta < 1$ and $\delta = 1$), in the absence of a commitment technology, the individual at the date of producing $t \in \{1, ..., n\}$ will underprovide effort relative to the optimal level from his perspective at date 0 ($e_1(0)$ rather than $e_0(0)$, see Section 2).[13]

Suppose now that, at date 0, P may enter a contractual relation with another individual (the monitor M) who is also time-inconsistent. More specifically, the contract between P and M can specify an effort to be exerted e^* by P whenever production is necessary and the dates at which P and M meet. Upon such meetings, M costlessly checks the levels of previous efforts by P provided these were exerted no more than y (≥ 1) dates before. Formally, we assume that if P exerts an effort $e < e^*$ at date t, then the probability that M detects P shirking in period $t + \tau$ is:[14]

$$\begin{cases} p(e^* - e) & \text{if} \quad \tau \leq y \\ 0 & \text{if} \quad \tau \geq y + 1 \end{cases}$$

[13] Given $\delta = 1$, the surplus of production for self-0 is the same independently of the date of production. In Section 3.1.3 we generalize this setting by studying the pure hyperbolic discounting case.

[14] Instead of a constant probability of detection that drops to zero after some periods, one could assume a function $p(\cdot)$ smoothly decreasing in τ. The results would not change significantly under this alternative formalization.

The time at which effort has to be exerted is unknown at date 0 (for simplicity, it is uniformly distributed between 1 and n). It is only learned by P at the beginning of the period in which production takes place and, more specifically, before meeting with M (if a meeting is scheduled for that period). Therefore, upon a meeting, parties can use this opportunity not only to check past effort but also to *renegotiate* prescribed effort levels for that date or for any future date.

Given that both individuals are time-inconsistent, if effort is required at date t, the *joint surplus* of P and M (including any frictionless interpersonal transfer) from their self-0 and self-t perspective are respectively:

$$\beta\,[e-\psi(e)] \qquad \text{and} \qquad \beta\,e-\psi(e)$$

which is exactly the same as in (1). It is important to notice that interpersonal contractual relations are no miracle cure to intrapersonal incentive problems: given that both individuals are time-inconsistent, they will renegotiate the effort level exactly as in the one-person game.[15] There is still a crucial difference between intrapersonal and interpersonal relations: for the latter to happen, some degree of *coordination* is necessary. This difference is summarized in the following assumption.

Assumption 4 *Interpersonal meetings cannot be organized on the spot, but have to be arranged in advance.*

In an intrapersonal game, "meetings with oneself" are not only possible at any moment but even unavoidable.[16] Therefore, self-renegotiation and self-checking of previous effort can and will be conducted at every date. By contrast, in an interpersonal game, it seems reasonable that coordination problems combined with limited time availability reduces the individual's ability to schedule meetings on the spot. In the literature on monitoring in hierarchies (see e.g. Calvo and Wellisz, 1978) having to monitor potentially many individuals reduces one's ability to monitor each one of them. In the same vein, Aghion

[15]We do not consider contracts between agents with different rates of time-preferences because we want the joint surplus to be as close as possible to the one-person case. This facilitates renegotiation at the date in which effort has to be exerted and therefore makes it the most difficult case for solving the intrapersonal problem. We assume that interpersonal transfers are frictionless exactly for the same reason.

[16]Some authors (e.g. Thaler and Shefrin (1981) or Loewenstein (1996)) argue that the individual is divided in several entities with conflicting objectives. Under this approach, one could defend Assumption 4 even in an intrapersonal game. We will limit our attention to a more traditional view (at least in economics), in which the individual does not have a conflict of goals *at a given date*.

and Tirole (1997) argue that having many agents can serve as a commitment device for a principal who wants to commit not to be "interventionist".

In this setting, a date-0 contract between the two individuals consists of three factors. First, a prescribed effort level e^* for the producer. Second, a transfer C from the producer to the monitor for being caught exerting less effort than the prescribed level. This penalty will depend on the agents' attitude towards risk. In the case of risk-neutrality and limited liability, individuals will set the highest possible penalty whenever the first-best cannot be achieved. Third, a set of dates at which P and M meet to check previous effort levels. Given Assumption 4, meetings have both costs and benefits from the individuals' self-0 perspective. On the one hand, the expectation of a future meeting keeps P on his toes, because he fears being caught at that time and having to pay the penalty. On the other hand, a meeting can also be used to renegotiate away previously set effort levels. Given these considerations, we have the following result.

Proposition 2 *For any probability of detection $p(e^* - e)$, interpersonal interactions alleviate intrapersonal incentive problems. If we restrict attention to deterministic meeting patterns, these should optimally be organized every $y + 1$ dates.*

Proof. Suppose that meetings are organized every x periods and recall that M observes with probability $p(e^* - e)$ a deviation from the prescribed effort level of P in the past y periods. Two cases must be analyzed separately.

If $x \leq y$, all past efforts can be observed by M. Renegotiation between P and M takes place with ex-ante probability $1/x$ (i.e. whenever the effort has to be exerted in the current period). Optimal effort is enforced with probability $(x-1)/x$, since P anticipates that otherwise he can be detected. Conditional on $x \leq y$, P optimally minimizes the probability of renegotiation and sets $x = y$.

If $x \geq y + 1$ shirking can go unnoticed. Effort can be enforced only if there is a meeting during one of the next y periods. This occurs with ex-ante probability y/x. With probability $1/x$ there is renegotiation. Last, with probability $(x - y - 1)/x$ there is no need to comply with the effort prescribed as there is no meeting in the following y periods. Conditional on $x \geq y + 1$, P optimally maximizes the probability of a meeting and sets $x = y + 1$.

The proof is completed by noting that $x = y + 1$ yields a higher utility than $x = y$: in both cases there is either high effort or renegotiation, and in the former the probability of renegotiation $(1/(y + 1))$ is lower than in the latter $(1/y)$. □

The intuition behind this result is the following. A meeting date is a period "lost" in terms of committing not to shirk: any effort level in excess of $e_1(0)$ to be exerted *at that date* will be renegotiated away. From this point of view, meetings should be organized as infrequently as possible. On the other hand, prescribing high effort levels can have an impact in the producer's decision only if he expects to be controlled within y periods. From this point of view, meetings should be organized so as to minimize the opportunity for P to "get away" with low effort. This means that, for each period without a meeting, the next meeting should take place no more than y periods later, thereby leading to our result. Under this optimal frequency of communication and with a uniform ex-ante distribution for the period at which effort has to be expended, ex-post control (and therefore high effort) will occur with probability $y/(y+1)$, and renegotiation (and low effort) will occur with probability $1/(y+1)$. The intertemporal joint welfare from the agents' perspective at the date of production t is therefore:

$$V^t = \frac{y}{y+1}\Big(\beta\Big[e_1(e^*) - p(e^* - e_1(e^*)) \times C\Big] - \psi(e_1(e^*))\Big) + \frac{1}{y+1}\Big(\beta\, e_1(0) - \psi(e_1(0))\Big),$$

where, compared to Section 2, e^* is the result of the same maximization problem $\mathcal{P}$ by self-0 and $e_1(e^*)$ is the same optimal self-1 effort conditional on the promise by self-0, except that the cost function $f(e^* - e)$ is now replaced by $p(e^* - e) \times C$.

To sum up, the key assumption of our analysis is that *interpersonal* relations (monitoring and renegotiation) require some amount of coordination. Consequently, when an individual learns that he has to exert effort and that no meeting is scheduled, he is happy to have this commitment device. Besides, at that time it is not possible to renegotiate the commitment away anymore, and therefore the individual has incentives to comply with the prescribed effort e^* for fear of being caught shirking later on with probability $p(e^* - e)$. Contracting with another agent is necessary because the constraint in organizing a meeting is not present in the case of an intrapersonal game: "self-meetings" are possible any time. Note also that avoiding renegotiation does not necessarily imply that all the effort specified in the contract is exerted: as shown in Proposition 1, self-0 may commit to a certain effort anticipating that this level will never be attained ($e_1(e^*) < e^*$). In that case, detection and interpersonal transfers occur in equilibrium with strictly positive probability. The shape of the cost function (here, the probability of detection $p(\cdot)$ and the maximal punishment C) is the crucial factor that determines which type of promises are optimal.

At this point two remarks are in order. First, in standard models with time-consistent

individuals, interpersonal relations are at best neutral in terms of incentive effects (e.g. under perfect contracting) and otherwise detrimental. With individual time-inconsistency, it is possible to give a new, "positive" role to interpersonal relations between self-interested individuals: they can partially solve intrapersonal incentive problems.[17] Second, note that if the monitor can only control the effort exerted in the current period ($y = 0$), then interpersonal interactions lose all commitment value: both parties will prefer to renegotiate rather than enforce any effort in excess of $e_1(0)$.

We conclude this section by considering two extensions of our framework, that stress the robustness of our conclusions. These extensions can however be skipped in a first reading of the paper.

3.1.2 Extension: random meetings

The reader might worry about the restriction to deterministic meetings previously imposed. In fact, we can show that *if we also allow random meeting patterns, then for any probability of detection the optimal interval between two periods with positive meeting probabilities is at least y and no more that $y + 1$.*

The idea behind this result is easy to grasp. First, the producer is constrained by a contract that specifies high effort only if he expects to be monitored within y periods. So, for each date with a meeting, the subsequent one should take place with positive probability no more than $y+1$ periods later. Similarly, for each date without a meeting, the subsequent one has to take place with positive probability no more than y periods later. Naturally, when the probability of a meeting is between 0 and 1, then the next monitoring should be delayed for no more than y or $y + 1$ periods.

The second issue when random meetings are possible is to determine the relative merits of monitoring often and monitoring with a high probability. Suppose that, at date 0, P and M have to agree between meeting once during the next $y+1$ periods with probability $\theta_1 + \theta_2$ or meeting twice (once with probability θ_1 and once with probability θ_2). The expected cost of getting together, which is determined by the probability of renegotiation, is the same in both cases: $(\theta_1 + \theta_2)/(y + 1)$. However, the producer's total probability of being controlled *at least once* (which is the relevant probability of being caught shirking) is $\theta_1 + \theta_2$ in the first case and $\theta_1 + (1 - \theta_1)\theta_2$ $(< \theta_1 + \theta_2)$ in the second one. That is,

[17] A similar conclusion is reached in Brocas and Carrillo (2000b) where competition between time-inconsistent agents for a single good may increase their welfare by alleviating their individual incentives to (inefficiently) rush into pleasant but unreasonable activities.

there are decreasing returns to frequent checks. Few meetings with high probability is more efficient for monitoring than many meetings with low probability because catching an agent more than once for the same shirking activity is useless. For this reason, it is optimal to keep as much distance as possible between the periods in which monitoring occurs (with, of course, the constraint of not exceeding $y+1$). Putting the two arguments together we obtain the above result.

3.1.3 Extension: increasing the individuals' willingness to renegotiate

In this section we provide another robustness check of the analysis. We return to deterministic meetings and consider a slightly different version of the game. We assume that the time at which effort has to be exerted (unknown at date 0) is learned by P and M at date 1. Effort can only take two values $e \in \{\underline{e}, \overline{e}\}$ (where $\underline{e} < \overline{e}$). Upon meeting, parties cannot change future meeting dates.[18] Last, individuals have pure hyperbolic discount functions. More precisely, from the perspective of self-t, date t' $(> t)$ is discounted at a rate $\frac{1}{1+b\,(t'-t)}$. According to this formalization, if the individual learns that effort has to be exerted in τ periods, his surplus is given by:

$$\frac{1}{1+b\,(\tau+1)}\,e - \frac{1}{1+b\,\tau}\,\psi(e),$$

in which case his desired effort level is determined by:

$$\psi'(\tilde{e}) = \frac{1+b\,\tau}{1+b\,(\tau+1)} \tag{6}$$

Given that the RHS of (6) is increasing in τ, there exists a time gap z such that the producer wishes to put effort $\overline{e}$ if production is going to take place in z or more periods and effort $\underline{e}$ otherwise (naturally, when $z = 1$ we are in the same case as in Section 3.1.1). Assume that, at date 0, agents only know that the time at which effort has to be exerted is uniformly distributed between z and n. In this setting, the contract between P and M should provide incentives to optimally exert effort $\overline{e}$ but renegotiation to effort $\underline{e}$ will occur whenever a meeting takes place at the time effort has to be expended or in the $z-1$ previous periods. It is then straightforward to see that *for any probability of detection $p(\cdot)$ and given hyperbolic discounting, deterministic meetings should optimally be organized every $z+y$ dates.*

[18] This is like a strong version of Assumption 4: it is not only impossible to organize new meetings but even to change the existing ones.

This result is, in a sense, a generalization of Proposition 2, with the same intuition. Meeting can be costly because an effort level $\overline{e}$ to be exerted at that date or in any of the $z-1$ subsequent periods will be renegotiated away. However, the threat of a control by M within the next y periods is the only way of inducing P to exert high effort. These two opposing incentives lead to the result. Under this optimal frequency of communication and with a uniform ex-ante distribution of times at which effort has to be expended, low effort never remains unnoticed: with probability $y/(z+y)$ there is ex-post control (and therefore effort $\overline{e}$ is enforced) and with probability $z/(z+y)$ there is an agreement for a downward renegotiation to effort $\underline{e}$.

3.2 Reputation cost of breaking a promise

Consider the following stylized buyer/seller adaptation of the model presented in Section 2. A time-inconsistent seller (he) can, at date $t=0$, promise to deliver a good at $t=2$. To produce the good by that date, he has to exert some effort e at date $t=1$ with an immediate cost $\psi(e)$. This effort can be thought of as determining the *quality* of the good, with higher effort implying higher quality.

At $t=0$, the seller can promise an effort/quality level that we shall call e^p (and not e^*, for reasons that will become clear shortly). This cheap-talk promise is observed by the buyer (she), but the effort e actually exerted is not. The buyer must also take an action. More precisely, she must choose at $t=1$ the technology that will be used to transform the good purchased from the seller.[19] Let us call this action of the buyer e^*. One can think of this choice as the buyer trying to adopt the technology that is "most compatible" with the quality of the good produced by the seller. There is no cost associated to the selection of a specific technology. However, the total ex-post surplus of the trade will depend on both the buyer's technology and the seller's product quality. Formally, it is given by:

$$h(e, e^* - e).$$

where the *total* derivative of $h(\cdot)$ is increasing in e (a higher quality input is valuable for the ex-post surplus) and, at the same time, its partial derivative is decreasing in $|e^* - e|$ (the best technology is the one which fits most closely with the quality of the input).

We do not explicitly model how this surplus is split between the buyer and the seller. Instead, we suppose that the seller's date-2 benefit of production is, independently of his

[19] Since the buyer is only active in two consecutive periods (1 and 2), it is irrelevant whether her preferences are time-inconsistent or not.

(cheap-talk) promise e^p, equal to $e - f(e^* - e)$ if $e < e^*$ and to e if $e \geq e^*$. Therefore, the buyer's ex-post surplus is given by:

$$g(e, e^* - e) = \begin{cases} h(e, e^* - e) - \Big(e - f(e^* - e)\Big) & \text{if } e < e^* \\ h(e, e^* - e) - e & \text{if } e \geq e^* \end{cases}$$

where, as for $h(\cdot)$, we assume that the total derivative of $g(\cdot)$ is increasing in e and its partial derivative is decreasing in $|e^* - e|$.

The above assumptions are meant to match exactly the formalization of Section 2. They are also natural, since they are based on the idea that the seller has a payoff which increases in the quality of the input he produces, but that he loses some of this surplus if this quality is inferior to the one the buyer "has counted on". In other words, a difference between e^* and e will thus involve a loss for the buyer as well as for the seller. Note that this loss does not come directly from a difference between the promise e^p and the effort e, but from the buyer's choice of technology e^*. It is only when the buyer is unable to infer with certainty the effort e that a loss will be incurred. The timing is summarized as follows:

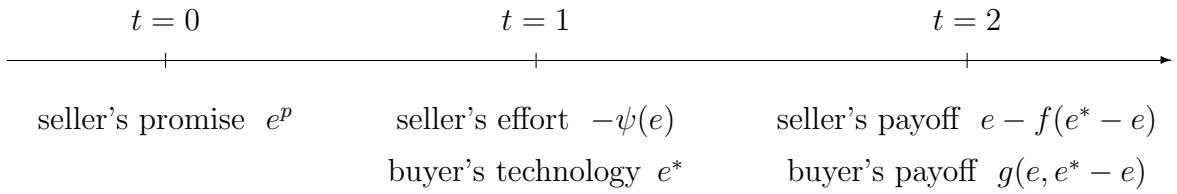


Figure 3. Timing of the reputation game.

Assume first that there is common knowledge of payoffs, and that the seller is time-consistent ($\beta = 1$). In this case, things are simple. The two parties play a Nash equilibrium at $t = 1$: the buyer optimally selects the technology that coincides with the anticipated effort of the seller ($e^* = e$), and the seller chooses his optimal effort level $e = e_0(0)$, where $\psi'(e_0(0)) = 1$. Promises at date 0 then play no role.

Consider now a second case, where common knowledge of payoffs is still assumed but where the seller is time-inconsistent ($\beta < 1$). Things are again simple. In any pure-strategy equilibrium, the buyer will select $e^* = e$ at date 1, and the seller will then choose an effort level $e = e_1(0)$, where $\psi'(e_1(0)) = \beta$. In this case, as of date 0, the seller would

like to commit his future self to a higher effort level, but he is unable to: whatever the date-0 promise e^p, the buyer will correctly understand the incentives of the seller when he takes his decision at date 1, so the promise will have no effect whatsoever.

We now introduce private information. Assume that the seller's cost of effort is:[20]

$$\gamma \cdot \psi(e),$$

where $\gamma \in \{\gamma_L, \gamma_H\}$ and $\gamma_H > \gamma_L > 0$. The seller privately knows the value of γ while the buyer has only a prior $p_H \equiv \Pr(\gamma = \gamma_H)$. If the seller is time-consistent, the date-0 promise can serve as a costless separating device. Call type i (with $i \in \{L, H\}$) the seller whose cost of effort is $\gamma_i \cdot \psi(e)$ and denote by $\hat{e}_i$ the optimal effort from his self-0 perspective, that is, the one that satisfies:

$$\gamma_i \cdot \psi'(\hat{e}_i) = 1. \tag{7}$$

It is in the interest of this seller to announce a promised effort level $e_i^p = \hat{e}_i$. Besides, since the type-i seller can be counted on to keep his promise ($e_i = e_i^p = \hat{e}_i$), the buyer will choose also a technology $e^* = \hat{e}_i$. Overall, promises *inform* the buyer about the agent's cost of effort, and therefore they are useful separating devices. However, the role of promises here is different from the one in Section 2 since they do not serve as *commitment* devices for future decisions.[21]

The most interesting situation arises when we simultaneously have private information about the cost of effort and time-inconsistency. Denote $\tilde{e}_i$ the optimal effort of type-i agent from his self-1 perspective. Formally, this effort satisfies:

$$\gamma_i \cdot \psi'(\tilde{e}_i) = \beta. \tag{8}$$

Recall that the optimal effort from his self-0 viewpoint is $\hat{e}_i > \tilde{e}_i$. Therefore, as in Section 2, the seller at date 0 would like to commit his self-1 incarnation to a higher effort level ($e_i^p > \tilde{e}_i$). While there is nothing type L can hope to achieve, there may be room for type H to pool with type L in order to induce a higher technology choice e^* from the buyer than in the separating case. This will induce self-1 of type H to work harder, but it may carry the cost of a broken promise. More specifically, we have the following result.

[20] The above case thus simply implied that $\gamma = 1$ with probability 1.

[21] In fact, promises are cheap talk and serve only as a separating device. It is therefore irrelevant which promises are announced as long as each type of seller announces a different promise and the buyer is able to know which promise corresponds to each type.

Proposition 3 *Under private information and time-inconsistency, there are two types of equilibria with promises:*

(i) *A separating equilibrium (promises as information devices), where type i's promise is* $e_i^p = \tilde{e}_i$*, his effort exerted is* $e_i = \tilde{e}_i$*, and the buyer chooses technology* $e^* = \tilde{e}_i$*.*

(ii) *A pooling equilibrium (promises as commitment devices), where both types of sellers announce a promise* $e_i^p = \tilde{e}_L$ *and type L exerts effort* $e_L = \tilde{e}_L$*. Type H may either fulfill his promise (*$e_H = \tilde{e}_L$*) in which case the buyer also chooses* $e^* = \tilde{e}_L$*, or fail short of it (*$e_H < \tilde{e}_L$*) in which case the buyer chooses* $e^* \in (e_H, e_L)$*.*

A sufficient condition for pooling being optimal is $\beta\,\gamma_H < \gamma_L$*.*

Proof. First, note that it is always optimal for type L to announce $e_L^p = \tilde{e}_L$ and exert an effort $e_L = \tilde{e}_L$.[22] If a separating equilibrium exists, then the buyer can infer the type of the agent from his promise (as in the time-consistent case). Sellers exert the optimal effort from their self-1 perspective $\tilde{e}_i$, which will also correspond to the buyer's selected technology. The payoff of type H from his self-0 perspective is then:

$$\beta\left[\tilde{e}_H - \gamma_H \cdot \psi(\tilde{e}_H)\right]. \tag{9}$$

Instead, in a pooling equilibrium, type H makes the same promise as type L, that is $e_H^p = \tilde{e}_L$. The buyer then chooses a technology e^* which solves:

$$\max_{e^*} \quad p_H \times g(e_H, e^* - e_H) + (1 - p_H) \times g(\tilde{e}_L, e^* - \tilde{e}_L) \tag{10}$$

where e_H is correctly anticipated. Simultaneously, type H solves:

$$\max_{e_H} \quad \beta\left[e_H - f(e^* - e_H)\right] - \psi(e_H).$$

From (10), the seller knows that any downward departure from $\tilde{e}_L$ will induce the buyer to choose a technology strictly between the two efforts. Two cases are then possible.

First, if the cost $f(e^*(e_H) - e_H)$ of underperforming is steep enough at 0 and above, it is optimal to fulfill promises in equilibrium: $e_H = \tilde{e}_L$ and $e^* = \tilde{e}_L$. The payoff of the high-cost seller from his self-0's perspective is then:

$$\beta\left[\tilde{e}_L - \gamma_H \cdot \psi(\tilde{e}_L)\right]. \tag{11}$$

[22] His problem is simple because we have assumed that he has a cost for *underperforming* ($e < e^*$) but not for *overperforming* ($e > e^*$) relative to the buyer's choice of technology. Consequently, the lower-cost seller makes the same choices and obtains the same payoff in the separating and in the pooling equilibrium. This simplification is not crucial for the result that time-inconsistency and private information provide a foundation for equilibrium broken promises.

Second, if the cost $f(e^*(e_H) - e_H)$ of underperforming is not that steep, then in equilibrium $e_H < \tilde{e}_L$ and, by (10), $e_H < e^* < \tilde{e}_L$. Type H fails to fulfill his promise (for which he pays a cost) and his payoff from the perspective of date 0 is:

$$\beta \left[e_H - f(e^*(e_H) - e_H) - \gamma_H \cdot \psi(e_H) \right]. \tag{12}$$

From the analysis above, it is clear that the value of p_H and the shapes of $f(\cdot)$, and $g(\cdot)$ will determine whether promises can serve as a commitment device and, if they do, whether they will be broken in equilibrium. However, recall that self-0 of type H would ideally want to implement an effort $\hat{e}_H > \tilde{e}_H$. Obviously $\tilde{e}_L > \tilde{e}_H$. Therefore, from (9) and (11), a sufficient condition for the seller to prefer a pooling rather than a separating equilibrium is $\tilde{e}_L < \hat{e}_H$. Given (7) and (8) this occurs when $\beta/\gamma_L < 1/\gamma_H$. □

The idea of an equilibrium with promises as a commitment device is the following. By pooling on the promise, the high-cost seller prevents the buyer from learning which individual will deliver the good at date 2. If the optimal effort of the type-L seller from his self-1 viewpoint is sufficiently close to the optimal effort of the type-H seller from his self-0 viewpoint, then the latter will mimic the former. This comes at no cost since the technology chosen by the buyer will correspond exactly to this effort level. However, the type-H seller can also decide to underprovide effort compared to his low-cost peer. The buyer naturally anticipates the departure but she is unsure about which seller is going to deliver the input, since both have announced the same target $\tilde{e}_L$. She then chooses a technology strictly in-between the two levels, with a corresponding cost for the agent who underperforms, but no cost for the one who overperforms.

As p_H increases, the buyer is more confident that she will face a high-cost seller. Then, the technology she chooses if $e_H \neq e_L$ becomes closer to e_H. This reduces the cost of breaking promises but, at the same time, it also decreases the value of the promise as a commitment device.[23] Note also that pooling can be excessively costly only when it induces the type-H seller to exert effort above his self-0 first-best level ($\tilde{e}_L > \hat{e}_H$). This is never the case if the intrapersonal conflict is sufficiently important (β low) and the two costs (γ_L and γ_H) are sufficiently close to one another.

Proposition 3 thus provides foundations for the results of Section 2, showing that time-inconsistent individuals are ready to make promises in order to commit future selves to alter their effort level. An equilibrium with promises is sustainable even though every

[23] In the extreme case $p_H = 1$, we are back to the situation with no private information in which promises are useless.

agent understands that future selves will (partially) renege on these promises if the cost of doing so is not too high. The cost of breaking promises here comes from one's *reputation*: just as in reputation models, the promise is made in order to keep the buyer uncertain about the seller type, i.e. to keep her believing that the seller might have a low effort cost. This leads her to revise upwards her technological decision, which is what even self-0 of the high-cost seller may prefer, despite the fact that a cost of breaking the promise may be incurred later on.[24] We finally illustrate our result with the following example.

Example 3: pooling and separating reputation equilibria. Consider the case where $f(e^* - e) = (e^* - e)$, $\psi(e) = e^2/2$, $g(e, e^* - e) = e - (e^* - e)^2/2$ and $e \in [0, 1]$. Besides, $p_H = p$, $\gamma_L = 1$ and $\gamma_H = 1/\alpha$ with $\alpha < 1$.

1. From (7) and (8), we have: $\hat{e}_H = \alpha$, $\hat{e}_L = 1$, $\tilde{e}_H = \alpha\beta$ and $\tilde{e}_L = \beta$. Given (9), if a separating equilibrium exists, the welfare of type H from his self-0 perspective is:

$$\beta \left[\alpha\beta - (\alpha\beta)^2/2\alpha\right] = \alpha\beta^2 (1 - \beta/2).$$

Note that a sufficient condition for pooling being optimal is $\tilde{e}_L < \hat{e}_H \Rightarrow \beta < \alpha$.

2. If a pooling equilibrium exists, then conditional on the effort e_H exerted by type H, the optimal technology adopted by the buyer solves:

$$\max_{e^*} \quad p \cdot [e_H - (e^* - e_H)^2/2] + (1 - p) \cdot [\beta - (e^* - e_H^2)/2].$$

Therefore, for all $e_H \leq \beta$, $e^* = p \cdot e_H + (1 - p) \cdot \beta$ and $(e^* - e_H) = (1 - p)(\beta - e_H)$.

3. In a pooling equilibrium, the optimal date-1 effort of type H (if $e_H < \beta$) solves:

$$\max_{e_H} \quad \beta[e_H - (e^* - e_H)] - e_H^2/2\alpha.$$

Therefore, $e_H = \alpha\beta(2 - p) < \tilde{e}_L$ if $\alpha(2 - p) < 1$ (unfulfilled promises) and $e_H = \beta = \tilde{e}_L$ if $\alpha(2 - p) \geq 1$ (fulfilled promises).

4. Suppose that $\alpha(2-p) \geq 1$. In a pooling equilibrium with fulfilled promises, the welfare of a type-H seller from his self-0 viewpoint is:

$$\beta \left[\beta - \beta^2/2\alpha\right] = \beta^2 (1 - \beta/2\alpha).$$

[24] As the reader might have noticed, in both sections 3.1 and 3.2 we need to introduce a second agent (monitor, buyer) in order for promises to play a commitment role. If we assume that agents have bounded memory (as in Benabou and Tirole (1999)), then it is possible to provide a reputational foundation for promises even in the absence of other individuals.

Hence, the necessary and sufficient condition for pooling being optimal is:

$$\beta^2 (1 - \beta/2\alpha) > \alpha\beta^2 (1 - \beta/2) \;\Rightarrow\; \beta < 2\alpha/1 + \alpha.$$

which, obviously, is a weaker condition than $\beta < \alpha$.

4 Foundations for promises based on contract incompleteness

In the earlier sections of this paper, we have analyzed the role of promises as a way to alleviate time-inconsistency problems rooted in individual preferences. These results are however also interesting because what they rely upon is a feature which goes beyond time-inconsistent preferences, namely commitment problems for individuals. While such problems naturally arise in the presence of individual 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. His inability to commit to exert his first-best effort level hurts him in terms of expected utility (see for example Jensen and Meckling (1976) as one example out of a large literature). Second, the *ratchet effect*, where the impossibility for the principal to commit not to take advantage of productive agents by subsequently raising their workload leads him to suffer from underprovision of effort by the agents early on (see for example Freixas et al. (1984) and Laffont and Tirole (1988)). Third, the *soft budget constraint* syndrome, where the inability of the principal to commit to terminate bad projects hurts creditors in situations where the threat of termination can deter bad entrepreneurs from asking for funds in the first place (see Dewatripont and Maskin, 1995).

The literature has investigated various ways out of these commitment problems. For example, in the ratchet effect and soft budget constraint literatures, having a principal with less information may reduce the commitment problem. As far as moral hazard is concerned, career concerns have been put forward since the work of Holmström (1999) as a mechanism that induces effort in the absence of explicit incentive schemes: the agent then works not to boost his current wage (which, at that time is already fixed), but in order to "impress the market" and thereby obtaining higher deferred compensation.

In the above circumstances, an alternative way to solve these commitment problems is to make promises as in Sections 2 and 3. Indeed, these commitment problems all imply de

facto the existence of two different "selves": "self-0", before the agent gets his contractual remuneration package, and "self-1", once the contract has been signed. And self-0 would like to commit self-1 to work harder than what self-1 will wish to. In the remainder of this section, we briefly highlight how we can reinterpret the two subsections of Section 3 in terms of contractual incompleteness instead of in terms of individual time-inconsistency.

4.1 Financial cost of breaking a promise

Consider the following reinterpretation of the model presented in Section 2. An agent is hired at $t = 0$ and receives a *fixed salary* w to produce a good for the Principal. Production takes place at $t = 1$. It requires a cost $\psi(e)$ to the agent to obtain an output e for the Principal (with $\psi' \geq 0$, $\psi'' \geq 0$, $\psi''' \geq 0$, $\psi(0) = 0$, and $\psi'(0) = 0$). The efficient level of effort is therefore $\hat{e}$, where $\psi'(\hat{e}) = 1$.

In a competitive labor market, agents are paid according to their productivity. However, if we assume that wages cannot be contingent on current productivity, then the agent has no incentives to exert effort at the production stage. The equilibrium therefore implies $e = 0$ and $w = 0 - \psi(0) = 0$. This is the standard moral hazard inefficiency when the agent cannot be made residual claimant for his effort. In other words, fixing wages before production yields formally the same type of inefficiency as having time-inconsistent preferences.

Suppose now that once the agent is hired he can promise to deliver an output e^* to the principal. If the principal observes that this output is not reached, the agent commits to give his salary w back to the principal.[25] Denote as before $p(e^* - e)$ the probability of observing that the target level has not been attained. Assume that $p(0) = 0$, $p'(0) = 0$, $p' \geq 0$, $p'' \geq 0$, and $p''' \leq 0$. For any given salary w and any promise e^*, the effort exerted at the production stage is:

$$\tilde{e}(e^*) = \arg\max_{e} \quad -\psi(e) - w \cdot p(e^* - e) \quad \Rightarrow \quad \psi'(\tilde{e}(e^*)) = w \cdot p'(e^* - \tilde{e}(e^*)). \tag{13}$$

Using the same techniques as in Proposition 1, note that:

$$\frac{\partial \tilde{e}}{\partial e^*} = \frac{w\, p''(e^* - \tilde{e})}{w\, p''(e^* - \tilde{e}) + \psi''(\tilde{e})} \quad \in (0, 1) \qquad \text{and} \qquad \frac{\partial^2 \tilde{e}}{\partial (e^*)^2} < 0.$$

Moreover, since $\psi'(0) = p'(0) = 0$, then $\tilde{e} \in (0, e^*)$ for all $e^* > 0$.

[25] One can think of this reimbursement as the maximum penalty for a cash constrained agent.

For any promise e^* and given a competitive labor market, principals compete à la Bertrand for agents (i.e. they get zero profit). Wages are therefore given by:

$$-w + \tilde{e} + w \cdot p(e^* - \tilde{e}) = 0 \quad \Rightarrow \quad w = \frac{\tilde{e}}{1 - p(e^* - \tilde{e})}. \tag{14}$$

The agent then sets the promise so as to maximize his production surplus:

$$\begin{aligned} e^* &= \arg\max_e \; w - \psi(\tilde{e}(e)) - w \cdot p(e - \tilde{e}(e)), \\ &= \arg\max_e \; \tilde{e}(e) - \psi(\tilde{e}(e)). \end{aligned}$$

The first-order condition, together with (13) and (14), implies:[26]

$$\frac{\partial \tilde{e}}{\partial e^*}\left[1 - \psi'(\tilde{e})\right] = 0 \quad \Rightarrow \quad \tilde{e} = \frac{1 - p(e^* - \tilde{e})}{p'(e^* - \tilde{e})},$$

which yields a unique solution for e^*. Overall, the agent will succeed in exerting his first-best effort level ($\psi'(\tilde{e}) = 1$) but, to achieve this, he will need to set an unrealistic promise ($e^* > \tilde{e}$). The cost of this promise (given by the probability of being detected) is fully recouped ex-ante via the wage. This last result is due to the competitive labor market assumption. In a more realistic setting with bargaining between parties, the agent would not achieve his first-best payoff and effort level. However, we would still get that due to contract incompleteness: (i) promises increase welfare, (ii) they are partially broken in equilibrium, and (iii) they entail a financial cost for the individual who makes and breaks them.

4.2 Reputation cost of breaking a promise

We can reinterpret the model of subsection 3.2 as follows. Instead of thinking of time-inconsistency as the reason behind the fact that incentives to exert effort for the seller are lower at $t = 2$ than at earlier periods, one can think of a traditional *holdup* problem. At the beginning of the game, the buyer would be ready to give the seller a payoff equal to e, but he cannot commit not to renegotiate this to βe once e has been chosen by the seller. Contract incompleteness enters in the following way: initially, the seller can obtain an additional fixed fee from the buyer, equal to $(1 - \beta)e$, if we assume that the seller has full bargaining power at that initial point and the buyer expects to earn this amount later on. However, since both parties understand that the seller will choose an effort level at $t = 1$ such that $\psi'(e) = \beta$, the seller in the end suffers from exactly the same commitment

[26] Note that the second-order condition holds by the concavity of $\tilde{e}(e^*)$.

problem as in subsection 3.2: he would like instead to be able to commit to choosing an effort level such that $\psi'(e) = 1$. In this setup, incomplete information and promises can help, exactly as with time-inconsistent preferences: if there is incomplete information about the cost of effort of the seller, there is the opportunity for high-cost sellers to pretend to be low-cost ones in order to induce the buyer to make a more ambitious technology choice e^*. This will have commitment value for the seller, but at a potential cost of $f(e^* - e)$ if the seller choice e is below this technology choice e^*. Whether this pooling behavior is profitable and whether it leads to unfulfilled promises depends on exactly the same conditions as in the Proposition 3.[27]

5 Conclusion

This paper has identified conditions under which promises, made by a time-consistent individual and which lead to a financial or reputation loss if broken, are (partially) broken. Two different foundations for the cost of broken promises have been considered. First, an endogenous financial loss arising from interpersonal monitoring and explicit contracting. Second, a reputation loss in the presence of incomplete information. Finally, we have argued that the above results are in fact pretty general: they remain valid when we replace time-inconsistent preferences by limits to contracting as the source of the commitment problem of the individual.

Further exploring the generality of our results would be an interesting avenue for future research. First, we have only sketched the multiperiod extension of the problem. Looking at the evolution of reputation in our context would certainly be interesting. Multitask extensions would also be natural topics for further study: Which tasks would be chosen for extending promises? How would the possibility of making promises affect the portfolio of tasks pursued by individuals? These issues are part of the general problem of organization design, in which for example the question of deadlines or timetables is naturally connected to this paper.

Finally, we feel that the connection between time-inconsistent preferences and limited contracting has been underresearched so far. Given their close parallel, as this paper illustrates, the literatures on these two topics would benefit from further cross-fertilization.

[27]Indeed, at $t = 2$, the seller will obtain a payoff of $\beta(e - f(e - e^*))$ if $e < e^*$, and at $t = 0$ he can extract another $(1 - \beta)(e - f(e - e^*))$.

References

1. Aghion, P. and J. Tirole (1997), "Formal and Real Authority in Organizations." *Journal of Political Economy*, 105, 1-29.

2. Ainslie, G. (1975), "Specious Reward: a Behavioral Theory of Impulsiveness and Impulse Control." *Psychological Bulletin*, 82, 463-496.

3. Ainslie, G. (1992), *Picoeconomics*, Cambridge University Press.

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

5. Benabou, R. and J. Tirole (2000), "Self-confidence: Intrapersonal Strategies." *mimeo*, Princeton and Toulouse.

6. Bleichrodt, H. and M. Johannesson (2000), "Time Preference for Health: a Test of Stationarity versus Decreasing Timing Aversion." forthcoming in the *Journal of Mathematical Psychology.*

7. Brocas, I. and J.D. Carrillo (1999), "Entrepreneurial Boldness and Excessive Investment." *CEPR D.P.2213.*

8. Brocas, I. and J.D. Carrillo (2000a), "The value of Information when Preferences are Dynamically Inconsistent." *European Economic Review*, 44, 1104-1115.

9. Brocas, I. and J.D. Carrillo (2000b), "Rush and Procrastination under Interdependent Activities." Forthcoming in *Journal of Risk and Uncertainty.*

10. Caillaud, B., D. Cohen and B. Jullien (1996), "Towards a Theory of Self-Restraint." *mimeo*, CEPREMAP, Paris.

11. Calvo, G. and S. Wellisz (1978), "Supervision, Loss of Control and the Optimal Size of the Firm." *Journal of Political Economy*, 86, 943-952.

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

13. Dewatripont, M. and E. Maskin (1995), "Credit and Efficiency in Centralized and Decentralized Economies." *Review of Economic Studies*, 62, 541-555.

14. Ellingsen, T. and M. Johannesson (2000), "Is there a Hold-up Problem?" *mimeo*, Stockholm.

15. Freixas, X., R. Guesnerie, and J. Tirole (1984), "Planning Under Incomplete Information and the Ratchet Effect." *Review of Economic Studies*, 52, 173-191.

16. Grossman, S. and O. Hart (1983), "An Analysis of the Principal-Agent Problem." *Econometrica*, 51(1), 7-45.

17. Holmström, B. (1999), "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies*, 66, 169-182.

18. Holmström, B. and D. Kreps (1995), "Notes on a Theory of Promises." *mimeo*, MIT.

19. Jensen, M. and W. Meckling (1976), "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure." *Journal of Financial Economics*, 3, 305-360.

20. Jovanovic, B. and D. Stolyarov (2000), "Ignorance is Bliss." *mimeo*, NYU.

21. Laffont, J.J. and J. Tirole (1988), "The Dynamics of Incentive Contracts." *Econometrica*, 56, 1153-1175.

22. Laibson, D.I. (1997), "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112, 443-477.

23. Loewenstein, G. (1996), "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65, 272-292.

24. Loewenstein, G. and D. Prelec (1992), "Anomalies in Intertemporal Choice: Evidence and an Interpretation." *Quarterly Journal of Economics*, 57, 573-598.

25. Mazur, J.E. (1987), "An Adjustment Procedure for Studying Delayed Reinforcement." in M. Commons, J. Mazur, J. Nevins, and H. Rachlin, eds., *Quantitative Analysis of Behavior: The Effect of Delay and of Intervening Events on Reinforcement Value*, Hillsdale: Ballinger.

26. O'Donoghue, T. and M. Rabin (1999), "Incentives for Procrastinators." *Quarterly Journal of Economics*, 114, 769-816.

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

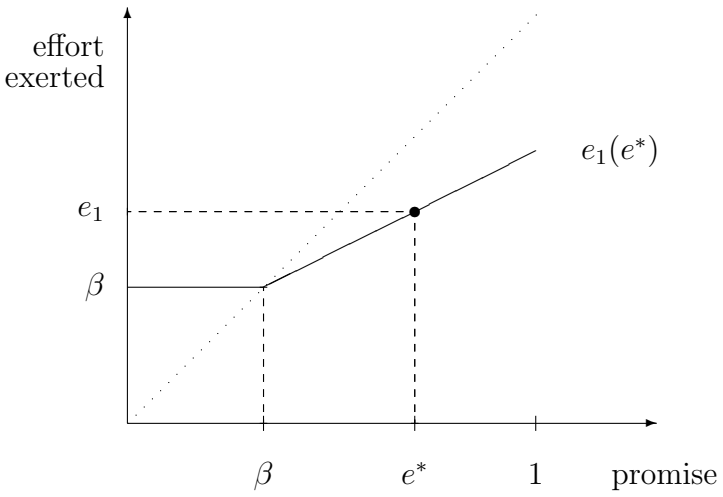
28. Rubinstein, A. (2000), "Is it "Economics and Psychology"?: the Case of Hyperbolic Discounting." *mimeo,* Tel Aviv and Princeton.

29. Strotz, R.H. (1956), "Myopia and Inconsistency in Dynamic Utility Maximisation." *Review of Economic Studies*, 23, 166-180.

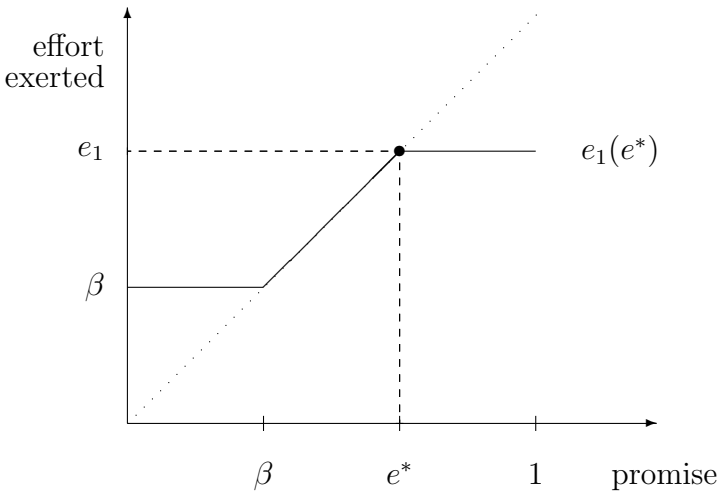
30. Thaler, R.H. (1981), "Some Empirical Evidence on Dynamic Inconsistency." *Economics Letters*, 201-207.

31. Thaler, R.H., and H.M. Shefrin (1981), "An Economic Theory of Self-control." *Journal of Political Economy*, 89, 392-406.

32. Womack, J.P., D.T. Jones and D. Ross (1990), *The Machine that Changed the World*, HarperPerennial, New York.



Example 1. Unfulfilled promises



Example 2. Fulfilled promises

Figure 2. Some examples.