The Brain as a Hierarchical Organization

By

Isabelle Brocas and Juan D. Carillo

April 2006

IEPR WORKING PAPER 06.48

INSTITUTE OF ECONOMIC POLICY RESEARCH

UNIVERSITY OF SOUTHERN CALIFORNIA

http://www.usc.edu/iepr

The Brain as a Hierarchical Organization *

Isabelle Brocas
USC and CEPR

Juan D. Carrillo
USC and CEPR

Abstract

We model the brain as a multi-agent organization. Based on recent neuroscience evidence, we assume that different systems of the brain have different time-horizons and different access to information. Introducing asymmetric information as a restriction on optimal choices generates endogenous constraints in decision-making. In this game played between brain systems, we show the optimality of a self-disciplining rule of the type "work more today if you want to consume more today" and discuss its behavioral implications for the distribution of consumption over the life-cycle. We also argue that our dual-system theory provides "micro-microfoundations" for discounting and offer testable implications that depart from traditional models with no conflict and exogenous discounting. Last, we analyze a variant in which the agent has salient incentives or biased motivations. The previous rule is then replaced by a simple, non-intrusive precept of the type "consume what you want, just don't abuse".

*We thank R. Bénabou, P. Glimcher, I. Palacios-Huerta, H. Shefrin and seminar participants at USC, Princeton, Columbia, Toulouse and Stanford SITE for comments and suggestions. Address for correspondence: Isabelle Brocas or Juan D. Carrillo, Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA - 90089-0253, e-mail: <brocas@usc.edu> or <juandc@usc.edu>.

"The heart has its reasons which reason knows nothing of"
(Blaise Pascal (1670), *Les Pensées*)

1 Introduction

In recent years, economics has experienced an inflow of refreshing ideas thanks to the addition of elements from behavioral psychology into formal models (see Rabin (1998) and Tirole (2002) for partial but insightful surveys). In these studies however, the brain has always been considered a black box. The time has come to open this box. For this enterprise, economists need the help of neuroscientists. As skeptics might object, this implies exploring unfamiliar territory and relying on evidence from a field that is still at its infancy. However, even if many fundamental questions about the functioning of the brain are still unanswered, there are also some well-grounded theories.[1] The challenge for economic theorists is to incorporate these theories into economic models in order to improve our understanding of human decision-making.

Our paper rests on two unorthodox assumptions about the individual's mind. First, there is an intrapersonal conflict of preferences within each period between two separate entities or systems of the brain. Second, there is asymmetric information between these two entities. Starting from these premises, we construct an orthodox dynamic model of consumption and time allocation which is solved with orthodox tools borrowed from mechanism design and economics of information.

Given the controversial nature of these assumptions, it seems natural to start with a brief review of the support provided by various disciplines to our hypotheses. The existence of internal tensions and asymmetries have long been discussed by classical authors in a wide range of fields, including Freud (1927) in clinical psychology and Smith (1759) in economics. Philosophical theories about split motivations are also recurrent. Pascal (1670) suggested the existence of private information between emotion and reason (see the introductory quote) and Hume (1739) a hierarchical order. Behavioral psychologists also adhere to the idea that treating individuals as coherent, fully informed entities is overly simplistic. Cognitive dissonance (Festinger, 1957), self-deception (Gur and Sackeim, 1979) and many other behaviors can best be understood under the acknowledgement that

[1] For interesting and clear summaries of some results in neuroscience that are relevant for economics, see Camerer, Loewenstein and Prelec (2004a, 2004b).

individuals are subject to conflicting goals, conflicting desires and conflicting beliefs. In fact, economics is probably the field most reluctant to accept the intra-period divided motivation and information asymmetry assumptions.[2] In our opinion, it stems from the belief that, despite their intuitive appeal, there is no solid theoretical or empirical support for these hypotheses.

This is where neuroscience enters the picture. Neuro-experimental techniques identify brain activity in choice processes, which can help theorists determine the key ingredients for modelling conflicts, constraints, tradeoffs and eventual decision-making at the brain level.[3] A brief look at the recent neuroscience literature reveals a remarkable parallel between the view of the brain held by neuroscientists and the view of organizations held by economists. For instance, some common topics of discussion in neuroscience circles include competition and control (Carter et al., 1998), monitoring and information processing (Botvinick et al., 2001), resource allocation (Montague et al., 2004) and self-management (Ochsner and Gross, 2005). The two ingredients of special interest for our paper, time-horizon and preferential access to information, have also received attention. McClure et al. (2004) show that the limbic system is associated with impulsive decisions yielding immediate gratification whereas the cortical system is activated in intertemporal decisions, hence the time-evaluation conflict between brain systems (see also Camerer et al. (2004a, section 5.1)). Studies by Berns et al. (1997) and Whalen et al. (1998) among others show that some brain regions are activated in response to novelty (anterior cingulate) and emotional stimuli such as fear (amygdala) even in the absence of awareness by subjects, hence the informational asymmetry between brain systems (see also Camerer et al. (2004a, section 4.5) for other examples). Time-evaluation conflict and restricted information constitute the cornerstones of our research.

Our model can be summarized as follows. The individual chooses a pleasant (consumption) and an unpleasant (labor) activity during two periods. At each date, its myopic / impulsive / limbic system (the "agent") is concerned solely with current utility whereas

[2] As developed below, inter-period conflicts of the self (hyperbolic discounting, temptation and self-control or any other form of dynamic inconsistency) are somewhat more accepted.

[3] Neuroscience provides a method to understand motivations that supplements revealed preferences and hypothetical questionnaires. This is nicely expressed by Freeman (1997, p.112) in his discussion of the different modes of perception: "One is an objective mode through observation of our conduct and its impact on others; the other is a private mode through awareness, which is verbally reported to others. These two modes can now be supplemented by direct observations of brain dynamics during the processes of perception. We are just beginning to realize the potential of this third mode for explaining features of consciousness previously inaccessible to us".

the far-sighted / cerebral / cortical system (the "principal") weights equally utility at all remaining dates. Each unit of labor translates into one unit of income. The individual can save and borrow in a perfect capital market so that only the intertemporal budget constraint needs to be satisfied. Consumption is desirable but its marginal value varies from period to period and is only known to the myopic system. Last, the cerebral system can impose on the impulsive system its desired choices. Given the informational asymmetry, the principal cannot reach its first-best level of consumption and labor. Instead, it proposes a menu of consumption and labor pairs where the level of both activities are positively linked, and lets the agent choose its preferred pair as a function of its valuation. In other words, we show the endogenous emergence of a self-disciplining, intrapersonal rule of behavior "if you want to consume more today, no problem except that you have to prove it by also working more today" (Proposition 2). Naturally, this rule cannot be implemented if labor opportunities are restricted at a certain date, which has an important implication: consumption over the life cycle will depend not only on total wealth accumulated over the life cycle but also on the source of wealth (labor vs. income shocks) and on the period-to-period access to labor (Corollary 1). Interestingly, moving from full information to asymmetric information in our dual-system model predicts similar changes in behavior than moving from no discounting to positive discounting in a standard model without conflict: the relative consumption between first and second period increases and the relative labor between first and second period decreases. Yet, some other predictions are different. In our setting with asymmetric information, choices depend not only on current valuations but also on how much the individual usually likes to consume that particular good. We conclude by arguing that our model provides some foundations for discounting (Proposition 3). In a second step, we assume –also based on neuroscience experiments– that the actions of the agent are determined by some biased, excessive willingness to consume. We show that, when the agent has such salient incentives, it may become optimal for the principal to change its intervention into a simpler, non-intrusive "you can do anything as long as you don't abuse" rule-of-thumb (Proposition 4).

There is a substantial literature that models intrapersonal conflicts, either with hyperbolic discounting of future returns (Strotz (1956), Laibson (1997) and others) or with some other form of self-control problems (Caillaud et al. (1999), Gul and Pesendorfer (2001) and others).[4] Some of these works have explicitly studied the effects of imperfect

[4]See Caillaud and Jullien (2000) for a review of different ways to model time-inconsistent preferences,

self-knowledge on individual decision-making (Carrillo and Mariotti (2000), Brocas and Carrillo (2004), Bénabou and Tirole (2002, 2004) and Amador et al. (2004)). Our work departs from this body of research in three important respects. First, we do not deal with inter-period conflicts of the self i.e., with the individual at date t having different preferences about optimal choices at $t+1$ than the individual at $t+1$ and trying to impose them. Instead, we are concerned with *intra-period conflicts of the self* i.e., with the brain struggling with conflicting current preferences over current behavior.[5] Second, the source of private information is not that news collected by or transmitted to the agent at date $t+1$ are unavailable at date t. Instead, we posit *informational asymmetry within each period*, with the "heart" in Pascal's words having a better access than the "reason" to the source of desires and motivations.[6] These two characteristics lead naturally to the analogy of the brain as a hierarchical organization whose members have incentive problems and conflicting goals. Third, we do not study the per-period propensity of the individual to engage in one activity (household consumption, nutritional therapy, etc.) that has opposite immediate and long run consequences (pleasure vs. lower savings, suffering vs. health improvement, etc.). Instead, we analyze *two intertemporally linked actions*: consumption and labor, dieting and exercising, etc. Thus, in our model, increased future consumption can be achieved not only vía higher savings (i.e., lower present consumption) but also vía higher present or future labor. These distinctions might seem innocuous to some readers. Yet, all three ingredients are essential to answer the questions we are interested in: Can self-disciplining rules that link activities within each period be optimal? When is it better to leave freedom to the impulsive system? Do the source of income and the distribution of labor opportunities over the life cycle affect the distribution of consumption?

There is another strand of research closer in spirit to our paper, where several entities with conflicting goals fight for supremacy. The seminal works by Thaler and Shefrin (1981) and Shefrin and Thaler (1988) are, to our knowledge, the first studies that divide the brain

Caplin and Leahy (2001) for the time-inconsistency effect generated by anticipatory feelings, Palacios-Huerta (2004) for the relation between anticipatory feelings and time-inconsistent choices, and Bénabou and Pycia (2002) for a discussion of the link between the different approaches.

[5] In other words, we do not play the game self-t (principal) vs. self-$t+1$ (agent) but rather the game self-t + self-$t+1$ (principal) vs. self-t (agent 1) and self-$t+1$ (agent 2). Note that, contrary to our paper, in models with inter-period conflicts of the self a central theme is the feasibility and optimality of *commitment devices* i.e., the individual's ability to lock future choices.

[6] To our knowledge, Bodner and Prelec (2003) is the only other study with intra-period asymmetric information. They focus on a different issue, namely how the "gut" who knows some information that cannot be introspected by the "mind" uses actions to signal preferences to himself.

into a forward-looking principal who can restrain the choices of its myopic agents. They explain the benefits of commitment devices such as mandatory pension plans and lump-sum bonus in promoting savings. These papers have been elegantly extended and further developed by Fudenberg and Levine (2004) and Loewenstein and O'Donoghue (2004). Fudenberg and Levine argue that the split-self approach explains several empirical regularities that depart from traditional theories, including dynamic preference reversals and the paradox of risk aversion in the large and in the small. Loewenstein and O'Donoghue show that this framework sets a parsimonious benchmark to study the optimal decision to exert willpower. Benhabib and Bisin (2004) and Bernheim and Rangel (2004) propose other types of dual-systems models. The first one studies consumption of an individual who can invoke either an automatic process (susceptible to temptation) or a control process (immune to temptation but which requires attention). The second one analyzes addiction under the assumption that the individual operates in either a 'cold mode' in which he selects his preferred alternative or a 'hot mode' in which choices may be suboptimal given preferences. All these works open the black-box of the brain. They provide invaluable insights on human behavior and important policy implications. As we will develop below, our paper addresses different questions. Perhaps more importantly, our approach is also different and, as we view it, complementary to theirs. The starting point in these papers is the existence of some exogenous cost (cost of self-control, cost of exerting willpower, cost of attention, cost of hot choices) that inevitably leads to tradeoffs (fewer resources but better allocation, costly thinking but optimal decision-making, higher current utility but increased likelihood of a future hot mode). Rather than a *cost*, our paper rests on asymmetric information, a *constraint* on optimal decision-making. By focusing on this cognitive inaccessibility to our preferences and motivations, we take a priori no position on the tradeoff or deviation from optimal behavior that is likely to occur in equilibrium. Last, no previous work has studied the effect of incentive salience and biased motivations in decision-making.

All in all, this paper provides a modest step towards an alternative method to study decision-making: experimental neuroscience provides evidence about the 'organizational structure' of the brain systems and microeconomic theory offers a methodology to solve the 'games' played among these brain systems. Just as the recent behavioral economics literature, this approach may help understand behaviors that are difficult to reconcile with traditional theories. More importantly, our longer term goals are to provide micro-

microeconomic foundations for ingredients in decision-making that have been traditionally considered as exogenous (discounting being just one example) and to revisit the individual choice paradigm, moving from a decision-theory to a game-theory, multi-agent approach.

2 Intra-period conflicts of the brain

We consider an individual who lives two periods, indexed by $t \in \{1, 2\}$. At each period, the individual decides his level of consumption c_t ($\geqslant 0$) and labor n_t ($\in [0, \bar{n}]$) or, equivalently, the amount of leisure $l_t = \bar{n} - n_t$. The instantaneous utility of the individual is given by the following simple equation:

$$U_t(c_t, n_t; \theta_t) = \theta_t \, u(c_t) - n_t$$

where $u' > 0$, $u'' < 0$, and θ_t captures the idea that the pleasure, willingness, need or urge to consume varies from period to period. Each θ_t (sometimes referred to as "valuation" or "type") is independently drawn from the same continuous distribution in $[\underline{\theta}, \overline{\theta}]$ with $\overline{\theta} > \underline{\theta} > 0$, strictly positive density $f(\theta_t)$ and c.d.f. $F(\theta_t)$ that satisfies the standard monotone hazard rate conditions ($\frac{d}{d\theta}\left[\frac{F(\theta)}{f(\theta)}\right] > 0$ and $\frac{d}{d\theta}\left[\frac{1-F(\theta)}{f(\theta)}\right] < 0$).

One novelty of our approach consists in modelling the brain of the individual as consisting of two separate entities. First, there is one principal (she) who is utilitarian and forward-looking. Second, there is an agent at each date t (he, from now on called agent-t) who is selfish and myopic. More specifically, agent-t maximizes his instantaneous utility $U_t(c_t, n_t; \theta_t)$ without any concern for past or future agents. By contrast, the principal does not derive utility by herself. She simply maximizes the sum of utilities of the agents in the remaining periods. As mentioned in the introduction, this intra-period conflict of the self has been suggested in many disciplines. Freud (1927) referred to "the Ego and the Id" and Thaler and Shefrin (1981) provided a first formalization in economics under a "Planner and Doer" label. We will adopt a more neutral "Principal and Agent" terminology borrowed from contract theory. Based on neuroscience findings, the principal represents the brain's cortical system, able to take into account long-term consequences of actions and to make intertemporal tradeoffs, whereas agent-t represents the brain's limbic system at date t, interested exclusively in immediate gratification (McClure et al. (2004)).[7]

[7] To our knowledge, Shefrin and Thaler (1988) was the first work to explicitly relate the planner and doer to the cortical and limbic systems, respectively.

As an aside, it is interesting to notice that McClure et al. (2004) argue that their experiment provides support for hyperbolic discounting (Camerer et al. (2004a) reach that same conclusion). Although the importance of their work in improving our understanding of intertemporal decision-making is unquestionable, we would like to qualify their claim. In our view, their experiment suggests the existence of an intra-period conflict whenever the individual chooses between immediate and delayed gratification: two systems of the brain, a myopic and a forward-looking, are simultaneously activated. This intra-period conflict is precisely what our model captures. We agree that it may very likely lead also to a preference reversal of the hyperbolic discounting type. However, to reach this second stage of the reasoning, we would need first to have more evidence concerning the rate of time preference of each system and then build a specific model that formalizes the claim.

In order to sharpen the contrast between principal and agent but, most importantly, to minimize the exogenous reasons for time-preference, we assume that the principal weights equally the utility of present and future agents. Formally, the intertemporal utility S_t of the principal from the perspective of date t is:

$$S_1 = U_1(c_1, n_1; \theta_1) + U_2(c_2, n_2; \theta_2) \quad \text{and} \quad S_2 = U_2(c_2, n_2; \theta_2)$$

In economic terms, each agent has a discount factor $\delta = 0$ and the principal has a discount factor $\delta' = 1$. In what follows, we will assume that, at each date t, agent-t selects consumption and labor (c_t and n_t). However, the principal can, at no cost, restrain the agents' feasible set of choices (even to the point of allowing only one pair, if she wants to). This formalization captures the idea that the impulsive, limbic system has easier access to actions whereas the cognitive, cortical system has the capacity to plan strategically (the planner/doer interpretation of Thaler and Shefrin (1981) and Fudenberg and Levine (2004)). It is also consistent with the evidence in neuroscience that some brain areas are involved in control whereas others are responsible for decision-making (Carter et al. (1998), Montague et al. (2004) and others).

For each unit of labor, the individual obtains one unit of income that can be consumed at any period. There is a perfect capital market where the individual can save and borrow at the exogenous, positive risk-free rate r. As a result, the individual only has to satisfy the intertemporal budget constraint:

$$c_1(1+r) + c_2 \leqslant n_1(1+r) + n_2$$

This formalization has an immediate but important difference with the standard model with one decision (e.g., consumption) and an exogenous (deterministic or stochastic) income stream: future consumption can be increased not only by increasing savings (i.e, reducing current consumption) but also by increasing current or future labor. In other words, there is scope for rules that "compensate" pleasant (consumption) with unpleasant (labor) activities at a given period. This will play a crucial role in the analysis.

2.1 Benchmark case: full information

As a benchmark for our analysis, we first suppose that the principal knows the willingness to consume of the agent at dates 1 and 2. Given that she can impose her desired levels of consumption and labor at each period, the preferences or even the "existence" of the different agents is irrelevant for the analysis. The program $\mathbf{P_1^o}$ that the principal solves is:

$$\begin{aligned} \mathbf{P_1^o}: \quad \max_{\{c_1,n_1,c_2,n_2\}} \quad & \theta_1\, u(c_1) - n_1 + \theta_2\, u(c_2) - n_2 \\ \text{s.t.} \quad & c_t(\theta_t) \geqslant 0, \quad n_t(\theta_t) \in [0,\bar{n}] \quad \forall\, t,\, \theta_t && \text{(F)} \\ & c_1(\theta_1)(1+r) + c_2(\theta_2) \leqslant n_1(\theta_1)(1+r) + n_2(\theta_2) && \text{(BB)} \end{aligned}$$

where (F) is a feasibility constraint on the values of c_t and n_t and (BB) is the intertemporal budget balance constraint. Our first preliminary result characterizes the solution to this problem.

Proposition 1 ***(Full information: first-best consumption / labor)***
The optimal consumption and labor pairs at dates 1 and 2 imposed by the principal are:

$$u'(c_1^o(\theta_1)) = \frac{1+r}{\theta_1} \quad \text{and} \quad n_1^o(\theta_1) = \bar{n}$$

$$u'(c_2^o(\theta_2)) = \frac{1}{\theta_2} \quad \text{and} \quad n_2^o(\theta_2) = (c_1^o(\theta_1) - \bar{n})\,(1+r) + c_2^o(\theta_2)$$

This proposition is straightforward.[8] Since labor enters linearly the agents' utility function and savings have a positive net return, it is optimal for a principal who weights equally the utility of both agents to concentrate as much labor as possible in the first period. Consumption at date t is proportional to agent-t's valuation θ_t and, ceteris paribus, it is higher in period 2 than in period 1 because of the above mentioned net return on

[8] The proof is trivial and thus omitted. Note that it is implicitly assumed that $\bar{n}$ is such that $c_1^o(\theta_1)(1+r) + c_2^o(\theta_2) \in (\bar{n}(1+r), \bar{n}(2+r))$ for all θ_1, θ_2. It is straightforward to extend the analysis to other cases with corner solutions.

savings. As $r \to 0$, the allocation of labor between periods becomes irrelevant and inter-period differences in consumption are solely determined by differences in valuation.

Note that consumption levels are determined only as a function of valuations and interest rates, and second period labor is then adjusted to meet the intertemporal budget balance constraint. In other words, there is no intra-period link between how much the individual should consume and how much the individual should work. Obviously, the result depends on some modelling assumptions (in particular, the quasi-linear utility function of agents). However, we adopt this formalization of preferences precisely because having no exogenous link between the variables within each period constitutes the most interesting benchmark of comparison.

2.2 The split mind: conflicts with asymmetric information

The analysis becomes more interesting when we introduce informational asymmetries within the brain. We assume that the principal can still impose her preferred levels of consumption and labor but only agent-t knows his valuation at date t. Asymmetric knowledge or, more specifically, the fact that some brain areas (including consciousness) may be unaware of information to which other brain areas respond is a well-established phenomenon in neuroscience (see Berns et al. (1997), Whalen et al. (1998) and the references there-in). In our context, it seems reasonable to posit that, if there is asymmetric information, then the decision-maker will know his instantaneous urges or willingness to consume better than the planner. However, to the best of our knowledge, there is currently no research that studies the access to information of these two systems.[9] Private knowledge is problematic for the principal: despite her ability to choose the allocation of labor and consumption between periods, her optimal decision depends on that information. Furthermore, given the conflict of preferences between the utilitarian forward-looking principal and the selfish myopic agent, the latter is unlikely to reveal for free his willingness to consume in this 'game of the mind'.

A digression. There are two main reasons why we may want to incorporate asymmetric information in a dual system model of the mind. First and obviously, because there is evidence, at least in other contexts, that individuals have restricted access to their own motives and beliefs. Second, because this assumption generates *endogenous constraints on optimal choices*. We would like to underscore the importance of this methodological

[9] We hope that this and other related models will catch the interest of neuroeconomists on this question.

contribution of the paper. As discussed in the introduction, splitting the individual into different entities that play a non-cooperative game is a modelling device that has already been employed in papers with a dual-system approach. The problem is that the definition of objectives, costs and choice sets of the different entities may affect which behavior can be rationalized. In this paper, we adopt a more agnostic approach: we assume that asymmetric information is the sole constraint faced by the principal, and then allow her to design *any* mechanism in order to impose on the agents her preferred choices. This methodology (borrowed from the mechanism design literature) is "neutral", in the sense that it does not presuppose a specific tradeoff and therefore it is difficult to anticipate what kind of deviations from optimal behavior are likely to occur. We can then assess the plausibility of the model by evaluating the empirical relevance of the behaviors it predicts.

Under private information, the principal solves two programs, one at each date. By the very nature of the problem, the principal deals with agent-1 and agent-2 sequentially, so we will solve the game by backward induction. At date 2, there is no conflict of preferences between the principal and agent-2 (formally, $S_2 \equiv U_2$), so the principal does not need to restrict the choice set of agent 2. Assuming that agent-1 has consumed and worked (c_1, n_1) and that the weak inequality (BB) must be satisfied, the optimal levels of consumption and labor that agent-2 freely selects at date 2 are, just as in section 2.1, given by:

$$u'(c_2^*(\theta_2)) = \frac{1}{\theta_2} \quad \text{and} \quad n_2^*(\theta_2) = (c_1 - n_1)(1+r) + c_2^*(\theta_2)$$

The program at date 1 is more interesting. In order to induce agent-1 to reveal his information, the principal must design an optimal incentive contract. To analyze this problem, we apply familiar contract theory techniques (see e.g. Fudenberg and Tirole (1991, ch.7)) to this unusual optimization program. More precisely, the principal restricts the options of agent-1 to a menu of pairs $\{(c_1(\theta_1), n_1(\theta_1))\}_{\theta_1=\underline{\theta}}^{\overline{\theta}}$, with as many pairs as there are potential valuations. Agent-1 is free to choose any of these pairs. The mechanism is conceived in a way that there is self-selection depending on the privately known valuation: when agent-1's willingness to consume is θ', then he picks the pair $(c_1(\theta'), n_1(\theta'))$ designed precisely for him. Applying the revelation principle, this direct mechanism achieves the

maximal (second-best) welfare of the principal if it solves the following program $\mathbf{P_1^*}$:[10]

$$\mathbf{P_1^*}: \max_{\{(c_1(\theta_1),n_1(\theta_1))\}} \quad S_1 = \int_{\underline{\theta}}^{\overline{\theta}} \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) + E_{\theta_2}\Big[\theta_2\, u(c_2^*(\theta_2)) - n_2^*(\theta_2)\Big]\, dF(\theta_1)$$

$$\text{s.t.} \quad \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) \geqslant \theta_1\, u(c_1(\tilde{\theta}_1)) - n_1(\tilde{\theta}_1) \qquad \forall\ \theta_1,\, \tilde{\theta}_1 \qquad (\text{IC}^*)$$

$$c_1(\theta_1) \geqslant 0, \quad n_1(\theta_1) \in [0, \bar{n}] \qquad (\text{F}_1)$$

Relative to $\mathbf{P_1^o}$, in the new program the principal maximizes expected welfare and must satisfy an incentive compatibility (IC*) constraint (i.e., a constraint which ensures that an agent-1 with valuation θ_1 weakly prefers to select the menu designed for him rather than the menu designed for someone with valuation $\tilde{\theta}_1 \neq \theta_1$). Note that the constraint (BB) is binding and embedded in the second period choices $(c_2^*(\theta_2), n_2^*(\theta_2))$. The solution to program $\mathbf{P_1^*}$ characterizes the second-best levels of consumption and labor at date-1 from the principal's viewpoint given the costs called for by asymmetric information.

Proposition 2 ***(Asymmetry in the mind: intraperiod link consumption / labor)*** *When only agents know their valuation, the principal offers to agent-1 the following menu* $\{(c_1^*(\theta_1), n_1^*(\theta_1))\}_{\theta_1=\underline{\theta}}^{\overline{\theta}}$ *of consumption and labor pairs:*

$$u'(c_1^*(\theta_1)) = \frac{1+r}{\theta_1 + r\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}$$

$$n_1^*(\theta_1) = \bar{n} - \left[\overline{\theta}\, u(c_1^*(\overline{\theta})) - \theta_1\, u(c_1^*(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} u(c_1^*(x))dx\right]$$

Agent-1 with type θ_1 *chooses the pair* $(c_1^*(\theta_1), n_1^*(\theta_1))$ *so that higher valuation implies more consumption* ($dc_1^*/d\theta_1 > 0$) *but also more labor* ($dn_1^*/d\theta_1 > 0$). *The principal allows agent-2 any pair of consumption and labor provided that it satisfies* (BB). *Agent-2 selects:*

$$u'(c_2^*(\theta_2)) = \frac{1}{\theta_2} \quad \text{and} \quad n_2^*(\theta_2) = \Big(c_1^*(\theta_1) - n_1^*(\theta_1)\Big)(1+r) + c_2^*(\theta_2)$$

The idea behind this proposition is quite intuitive. Ideally, the utilitarian and forward-looking principal would like agent-1 to consume $c_1^o(\theta_1)$ and work as much as possible to save for future consumption, as described in Proposition 1. However, if asked directly,

[10] Contrary to most contract theory problems, this program has no participation constraint. In other words, agent-1 can pick any consumption/labor pair he wishes (the one designed for him or any other) but he cannot refuse to choose among these pairs, take an outside option, and achieve a minimal utility level. Note, however, that the bounds $c_1 \geqslant 0$ and $n_1 \leqslant \bar{n}$ play a related role in constraining the minimum consumption and maximum labor that can be imposed on agent-1.

the myopic and selfish agent-1 is going to overstate his consumption needs. To solve this dilemma, one possibility available to the principal consists in forcing agent-1 to provide the levels of consumption and labor that maximize her expected welfare. Formally, $(\tilde{c}_1, \tilde{n}_1) \in \arg\max_{(c_1,n_1)} \int_{\underline{\theta}}^{\overline{\theta}} S_1(c_1, n_1, c_2^*, n_2^*; \theta_1) dF(\theta_1) \Rightarrow u'(\tilde{c}_1) = (1+r)/E[\theta_1]$ and $\tilde{n}_1 = \bar{n}$. Proposition 2 shows that the optimal labor and consumption plans are very different from that. Intuitively, in order to induce agent-1 not to consume excessively, the principal proposes to him the following rule: "Tell me what is your willingness to consume. The higher the value you say, the higher will be the quantity I will allow you to consume but the higher will be the amount of work I will ask you to provide in exchange". Asking more work in exchange of more consumption is the best mechanism the principal can use to counter agent-1's lack of concern for the future.

It is essential to realize that the positive relation between the intertemporal levels of consumption and labor ("work more in your lifetime if you want to consume more in your lifetime" or, formally, the correlation between $c_1^* + c_2^*$ and $n_1^* + n_2^*$) *is not* a result but, instead, a direct consequence of the (BB) constraint imposed in our model. By contrast, the self-disciplining rule of "work more today if you want to consume more today" *is* a main result of the dual system model with asymmetric information. It is neither first-best (as we know from Proposition 1) nor an ad-hoc, externally imposed restriction. Instead, it emerges as the internal, self-imposed, endogenously optimal second-best rule designed by the cerebral side of the individual to counter the tendency of the impulsive side to overconsume. Stated differently, the model provides foundations for self-imposed behaviors such as "I will spend the week-end in Palm Springs only if I finish the paper by Friday" or "I watch the soccer game but then I work on the referee report" based exclusively on informational asymmetries within the individual.[11] Naturally, we could have obtained similar results by introducing a parameter of "guilt" in the utility function that kicks-in when only pleasant activities are undertaken. However, we feel that deriving this behavior from first principles is more satisfactory than imposing an intuitive but still somewhat ad-hoc cost.[12]

[11] Note that, in these examples, the unit of time is defined somewhat loosely (one week, one day). If we set shorter and shorter periods, then it becomes impossible to perform two activities at one date, and we are back to problems of dynamic choices with interperiod conflicts of the self (we thank Roland Bénabou for this remark).

[12] The reader might argue that sometimes we do experience guilt, a sentiment that in our model can not be felt in equilibrium. If we introduced some stochastic element in our model, we could keep the essence of our results and have agents suffering from regret or guilt depending on the ex-post resolution of

Note also that our self-imposed rule rationalizes narrow bracketing of the type documented by Camerer et al. (1997) in their study of labor supply choice by New York City cabdrivers. Our model predicts that the cortical system designs for the limbic system a rule that links positively daily earnings and daily amount of leisure, that is, it predicts a negative elasticity of wages and hours of work. Our paper however does not account for short term targeting, a second (and perhaps more controversial) finding in that study.

A comment on contract enforcement. The principal-agent literature relies either on self-enforcing contracts (e.g., through reputation) or enforceability by a third party (e.g., a court).[13] If we believe in a pure contracting interpretation of our model (which seems a bit too literal), then enforcement is indeed a concern: both activities can hardly be undertaken simultaneously, so what prevents the principal from reneging on her promise of consumption once labor has been carried out?[14] It is difficult to think of a dynamic self-enforcement mechanism given the agents' myopia. We can hypothesize the existence of another brain system whose role is to ensure that the cortical system keeps her end of the deal. This would be consistent with the existing evidence in neuroscience that some brain areas (such as the anterior cingular cortex) are involved in monitoring, error detection and control of activities by other systems (Carter et al. (1998), Botvinick et al. (2001)). However, there is currently no evidence (for or against) this hypothesis in our specific context.

Another interesting result follows from the previous analysis.

Corollary 1 ***(Consumption over the life cycle)***
The distribution of consumption over the life cycle depends not only on the total wealth accumulated over the life cycle but also on the source of wealth and the period-to-period access to labor.

In order to implement the welfare maximizing rule of consumption and labor described in Proposition 2, the upper bound on the amount of labor that the individual is able to provide must be sufficiently high.[15] In other words, an individual cannot compensate a limited access to current labor with the equivalent amount of income derived from another

uncertainty.

[13] In some contract theory problems, enforcement is a delicate issue that has raised controversies (see e.g. Laffont and Martimort (1997) on collusion and the debates it generated).

[14] We thank Douglas Bernheim for raising this issue.

[15] Formally, $n_1^*(\underline{\theta}) \geqslant 0 \Leftrightarrow \bar{n} > \overline{\theta}\, u(c_1^*(\overline{\theta})) - \underline{\theta}\, u(c_1^*(\underline{\theta})) - \int_{\underline{\theta}}^{\overline{\theta}} u(c_1^*(\theta))d\theta$ (see the Appendix for details).

source, including past labor, future labor and income shocks. The intuition relies again on the self-disciplining rule "if you want to consume, then work". The principal needs proof of agent-1's consumption needs. To obtain it, she has to ask in exchange something which is *costly for him*: current labor. Thus, for example, if agent-1 is not willing to work as much as $n_1^*(\theta')$, then it reveals that his valuation is lower than θ' and therefore that he should consume less than $c_1^*(\theta')$. What if an agent-1 with valuation θ' wants but cannot work the required amount ($\bar{n} < n_1^*(\theta')$)? Unfortunately, the principal cannot ask for another type of income in exchange: since agent-1 does not care about past or future consumption, he is willing to sacrifice any of this income (endowment, income inherited from past labor, income borrowed against future labor, etc.) for extra consumption, independently of his true valuation.[16] Overall, a restricted access to labor in a given period will limit the ability of the forward-looking entity to extract information about the willingness to consume by the myopic entity. This, in turn, will constrain the amount of consumption that can be granted in that period.

Note that this corollary has implications for a long-standing puzzle in theories of consumption over the life-cycle, namely the tendency of households to engage in insufficient consumption smoothing. Our theory suggests that, controlling for the amount of wealth accumulated during the life cycle and even with perfect capital markets, an individual will consume more when he is active on the labor market than when he is not. Moreover, consumption behavior will vary depending on the source of wealth (labor vs. unanticipated income shocks).

2.3 The determinants of time-preference

The reader familiar with contract theory might argue that imposing a monotone relation between the observable variables is quite common in optimization problems with asymmetric information. In this section, we make one step further in our analysis. We perform some comparative statics on the nature and extent of the distortion as a function of the information asymmetry between the myopic and forward-looking entities, which lead to some further interpretations.

Suppose that the agents' valuation θ_t may be drawn from two distributions, $F(\theta_t)$ or $G(\theta_t)$, with $G(\theta_t)$ being "more favorable" than $F(\theta_t)$ on $[\underline{\theta}, \overline{\theta}]$. Formally, we assume

[16] The effect is mitigated (but does not disappear) if each agent has some concern about the welfare of past or future agents.

that their density functions $f(\theta_t)$ and $g(\theta_t)$ satisfy the monotone likelihood ratio property (MLRP):[17]

$$\left(\frac{g(\theta_t)}{f(\theta_t)}\right)' > 0 \qquad \forall\ \theta_t \in [\underline{\theta}, \overline{\theta}]$$

Intuitively, this condition implies that the agent is more likely to have a high valuation and less likely to have a low valuation for a good drawn from G than for a good drawn from F. Denote by $(c_t^{**}(\theta_t), n_t^{**}(\theta_t))$ the analogue of $(c_t^*(\theta_t), n_t^*(\theta_t))$ when θ_t is drawn from G rather than F. We have the following result.

Proposition 3 ***(The endogenous rate of time-preference)***

(i) For any distribution function $F(\cdot)$ and any increasing and concave utility $u(\cdot)$:

$c_1^*(\theta_1) > c_1^o(\theta_1)\ \forall\ \theta_1,\ \ n_1^*(\theta_1) \leqslant n_1^o(\theta_1)\ \forall\ \theta_1$ *and* $U_1^*(\theta_1) > U_1^o(\theta_1)\ \forall\ \theta_1$.

$c_2^*(\theta_2) = c_2^o(\theta_2)\ \forall\ \theta_2,\ \ n_2^*(\theta_2) > n_2^o(\theta_2)\ \forall\ \theta_2$ *and* $U_2^*(\theta_2) < U_2^o(\theta_2)\ \forall\ \theta_2$.

(ii) Moreover, if $u(\cdot)$ is in the CRRA class of utility functions, then for all θ_1 and $F(\cdot)$: $\dfrac{dc_1^*}{d\theta_1} > \dfrac{dc_1^o}{d\theta_1}\ (>0)$, $\dfrac{dn_1^*}{d\theta_1} > \dfrac{dn_1^o}{d\theta_1}\ (=0)$ *and there exists $\bar{r}$ s.t.* $r \gtrless \bar{r} \Leftrightarrow \dfrac{dU_1^*}{d\theta_1} \gtrless \dfrac{dU_1^o}{d\theta_1}$.

(iii) Last, and again for any $u(\cdot)$, if G is more favorable than F then:

$c_1^*(\theta_1) > c_1^{**}(\theta_1)\ (> c_1^o(\theta_1))$ *and* $\left.\frac{dn_1^{**}}{dc_1^{**}}\right|_{\theta_1} > \left.\frac{dn_1^*}{dc_1^*}\right|_{\theta_1}\ \left(> \left.\frac{dn_1^o}{dc_1^o}\right|_{\theta_1}\right)$ *for all θ_1.*

It comes at no surprise that agent-1 enjoys a higher utility under asymmetric than under full information: a superior knowledge of his valuation must necessarily benefit him, otherwise he would have no incentives to keep it private. The interesting question is to determine in which dimension(s) he benefits. As part (i) demonstrates, for all interior valuations, agent-1 both consumes more and works less than under full information. The reason is that, under asymmetric information, the principal must propose a high-powered incentive scheme to agent-1 (work according to what you want to consume). This translates into an inefficiently low level of first-period labor. Given this work shortage, the principal could reduce also first-period consumption in order to satisfy the intertemporal budget constraint. This would decrease the "rents" obtained by the agent (that is, the extra utility enjoyed due to the existence of private information), but also the "efficiency" of the intrapersonal contract. Instead, we show that the principal allows also extra consumption and compensates both the excessive consumption and the insufficient labor in the first period with an increased amount of labor in the second period. What makes this alternative optimal is the absence of a conflict of interests between the principal and

[17] MLRP is a standard condition in the comparison of distributions. It implies first-order stochastic dominance and hazard rate dominance: $G(\theta) < F(\theta)$, $\frac{g(\theta)}{G(\theta)} > \frac{f(\theta)}{F(\theta)}$ and $\frac{g(\theta)}{1-G(\theta)} < \frac{f(\theta)}{1-F(\theta)}$ for all $\theta \in (\underline{\theta}, \overline{\theta})$.

agent-2: the former does not need to grant extra rents to the latter in order to elicit higher second period labor. Summing up, it was obvious that agent-1 would benefit from private information. What seems more striking is that, for all valuations, the benefit always translates both into more consumption and less labor. Pushing forward this analysis, part (ii) states that, as the willingness to consume increases, the upward distortion in consumption becomes more pronounced and the downward distortion in labor becomes less pronounced. Therefore, the major benefit from information asymmetry takes two different forms depending on agent-1's willingness to consume: increased consumption for a high-valuation agent-1 and reduced labor for a low-valuation agent-1. Last, part (iii) shows that not only the realized valuation matters for consumption and labor but also whether it is a good that the agent usually likes it a lot or not. A consumption in excess of c_1^o is granted to agent-1 to avoid over-representation of the true willingness to consume. As high willingness becomes more predictable (because valuations are drawn from more favorable distributions), the principal is less willing to give a premium for information.[18]

We can also provide a more interesting behavioral interpretation of these results. Recall that choices under full information are analogous to those made by an individual with no intraperiod conflict and a discount factor of 1. Therefore, the difference between the consumption ratios under asymmetric and full information (c_1^*/c_2^* vs. c_1^o/c_2^o) identifies an endogenously determined "rate of time-preference" or "degree of impatience". The same holds for labor. In other words, these differences capture how many resources are shifted from period 2 to period 1 due to asymmetric information. According to this definition, part (i) shows the existence of a positive rate of time-preference both in consumption and labor ($c_1^*/c_2^* > c_1^o/c_2^o$ and $n_1^*/n_2^* < n_1^o/n_2^o$). Also, part (ii) demonstrates that agent-1's main benefit takes two different forms depending on the agent-1's marginal value of consumption: a boost in consumption when θ_1 is high and a reduction in labor when θ_1 is low. Simple calculations demonstrate that these two results coincide with those obtained if, instead of endogenous and a consequence of informational asymmetries, we assumed no intrapersonal conflict and an exogenous discounting of future returns ($\delta < 1$). In that respect, we can argue that our model provides foundations for intertemporal discounting. A natural issue is then to determine whether we can discriminate between the two alternatives. Part (iii) provides testable predictions that depart from standard theories of discounting: given

[18]There is a similar effect in the contract theory literature, where firm efficiency, worker ability and product quality is more valuable (i.e., implies not only higher payoff but also higher informational rents) the lower the efficiency, ability and quality in the pool of firms, employees and manufacturers.

identical current valuation for two different goods, an agent will exhibit lowest consumption for the good that he usually enjoys most. Moreover, each extra unit of consumption will "cost" him more labor in exchange. This result relates to the findings of Loewenstein et al. (2001), according to which, the intertemporal preference rate of an individual differs across activities.

Another issue of interest is to determine which type of agent-1 benefits more from the asymmetry of information or, stated different, which agent is more difficult to keep under control. By construction, agent-1's total utility increases with his valuation both under complete and incomplete information ($dU_1^o/d\theta_1 > 0$ and $dU_1^*/d\theta_1 > 0$). As the return on savings increases, the opportunity cost of current consumption also increases, and so does the principal's willingness to indulge extra utility to agent-1 in exchange of a truthful revelation of θ_1. As a result, for high interest rates, agents with strong valuations reap the biggest benefits of information asymmetries whereas for low interest rates, agents with weak valuations enjoy the greatest rents.

Last, recall that the ability to withhold information is beneficial for agent-1. Since it is detrimental for a principal who is solely (and equally) concerned about the utility of both agents, then it must necessarily be detrimental for agent-2.[19] A conclusion follows directly from this result.

Corollary 2 ***(The relative weight of utilities)***
Under complete information, the utility of agent-1 is lower than that of an agent-2 with the same valuation. This conclusion does not necessarily hold under asymmetric information.

The positive net return on savings combined to the equal concern of the principal for the welfare of both agents implies that, *ceteris paribus*, agent-1 should ideally consume less, work more and therefore enjoy a lower overall utility than agent-2. Formally, $U_1^o(\theta) < U_2^o(\theta)$ for all θ. Under asymmetric information and due to the extra utility needed in order to elicit agent-1's true valuation, the balance is restored and may even be inverted: agent-1 consumes more than agent-2, and the realized valuations together with the other parameters of the model determine whether he also works less. Formally, $U_1^*(\theta) \gtrless U_2^*(\theta)$.

Reversing the logic, one may conjecture that the system of the brain with long-term concerns is "designed" to weight equally utility at every period. This way, future agents

[19] It would be interesting to extend our results to a more general model with T (> 2) periods. Our conjecture is that asymmetric information would affect positively the utility of the first s ($\geqslant 1$) agents and negatively the utility of the last $T - s$ ($\geqslant 1$) ones, with s depending on the parameters of the model.

are favored and the unavoidable extra benefit enjoyed by the current agent due to private information is countered. Needless to say, the argument is speculative. In particular, there is evidence that the cortical system is capable of making intertemporal tradeoffs but there is no conclusive indication whether it discounts future returns, let alone which purpose such discounting would serve. In any case, this problem translates to the brain level an interesting question long discussed in welfare economics: should a social planner (our principal) discount less heavily the future than individuals in the population (our agents)?[20]

2.4 A note on the applicability of this theory

The model presented in this section has three key ingredients. First, asymmetric information and conflict of interests between myopic agents and forward-looking principal. Second, two activities that affect welfare at each period. Third, an intertemporal link between some of these activities. Since, all the other features of the model can be relaxed, the scope of applicability of our theory goes beyond the consumption and labor example in which we have focused so far.

In particular, our methodology and results immediately extend to situations where (i) both activities yield positive utility and/or (ii) the intertemporal link affects only one activity and/or (iii) the intertemporal link imposes a weaker relation than strict budget balance. One can think of (i) as a situation where an individual with a fixed budget decides to allocate his monthly spendings between several pleasurable activities (e.g., movies in the evening and concerts at night). If the myopic system has private information about the current value of one or both activities, the second-best rule imposed by the forward looking system is to link negatively not only the monthly (as imposed by the budget constraint) but even the daily relative consumption of the activities ("it's either a movie or a concert today, you choose"). According to (ii), the same conclusion applies even if concerts are free and therefore do not affect the budget constraint. Only by restraining the agent's ability to go to concerts can the principal obtain information about his true willingness to go to the movies. The mechanism is suboptimal, since free concerts should be consumed every day. However, it allows a more efficient allocation of the budget between movies at different dates.[21] In that respect, our theory provides a rationale for self-handicapping or,

[20] For a recent perspective on this problem, see Caplin and Leahy (2004).

[21] It is easy to construct an example with two activities and a binary consumption of each (0 or 1) where the equilibrium pair of consumptions is $\{1, 0\}$ or $\{0, 1\}$ at each date but never $\{0, 0\}$ or $\{1, 1\}$ even if one

more precisely, for 'self-inflicted punishments'. Last, property (iii) implies for example that an individual can compensate the ingestion of products high in cholesterol with current or future exercise, but he can also choose to sacrifice part of his health.

3 Wanting, liking and the importance of visceral factors

3.1 Some evidence of incentive salience

A current strand of the neuroscience literature pioneered by Berridge (2003) argues that behavior is not always motivated by the pursuit of pleasure. His research on consumer decision-making shows that subliminal stimuli can alter manifested choices without affecting the expected pleasure derived from the commodities. Starting from this evidence, the author draws a distinction between the "liking" system –responsible for the feeling of pleasure and pain–, and the "wanting" system –responsible for the motivation or incentive to seek pleasure and avoid pain–. Contrary to prior theories, Berridge claims that wanting and liking are mediated by two distinct brain systems. A series of laboratory experiments demonstrates that intervention in the mesolimbic dopamine system can enhance the willingness of rats to work for food (the wanting system) without affecting the pleasure of eating it (the liking system).[22]

More generally, there has been a growing interest among scholars across disciplines in studying the interplay between affection and cognition. It has been argued that the affective system can help (Damasio (1994) and others), constrain (Elster (2004) and others) or prevent (Baumeister (2003) and others) the cognitive system from engaging in rational decision-making. Loewenstein (1996) provides a mathematical representation of the effect of visceral factors such as emotions (anger, fear) or drives (hunger, sexual arousal) on decision-making. The paper argues that these states result in discrepancies between optimal and realized (what he calls out-of-control) behavior.

Although Berridge's incentive salience and Loewenstein's out-of-control theories are different in nature and have different implications, they share an important feature. Both emphasize the existence of salient and biased motivations that preclude or at least constrain welfare maximizing decisions. In this section, we incorporate a stylized version of

activity is free and always yields positive utility.

[22] Evidence in animals is an extremely weak indicator of human behavior and thus should be heavily discounted. This research is still worth mentioning given the obvious impossibility to perform pharmacological manipulations in humans.

this dichotomy between liking vs. wanting or optimization vs. visceral influences in our dual system model of the brain. More precisely, we assume that the true instantaneous payoff of the agent at date t is, just as before, given by $U_t(c_t, n_t; \theta_t) = \theta_t\, u(c_t) - n_t$. This utility function captures the liking part of the individual, or how consumption and labor does affect welfare. However, what motivates agent-t to consume and work at date t is $V_t(c_t, n_t; \theta_t) = \alpha\, \theta_t\, u(c_t) - n_t$, where α (> 1) is a known parameter. This other utility function captures the wanting part of the individual, or how the visceral factors influence perceived welfare and behavior. So, an individual with a utility of consumption proportional to θ_t but who is motivated by $\alpha\, \theta_t$ will be tempted to consume "excessively".

As in section 2, the benevolent principal maximizes the welfare of both agents ($U_1 + U_2$) whereas the myopic agents are concerned exclusively by their current utility. Unlike before, agent-t's motivation to work and consume is given by an inaccurate perceived utility ($V_t \neq U_t$). Under complete information, selfishness and biases in the agents' utility is irrelevant. The principal can impose her optimal pairs $(c_t^o(\theta_t), n_t^o(\theta_t))$ as described in Proposition 1. Under incomplete information and given that the principal now faces a conflict with both agents, she must design a revelation mechanism with each of them. Interestingly and as we develop below, the options offered to agent-1 and agent-2 are very different.

3.2 Biased motivations at date 2

Let us first analyze the game between the principal and agent-2. Suppose that agent-1 has consumed $\hat{c}_1(\theta_1)$ and worked $\hat{n}_1(\theta_1)$, and denote by $k(\theta_1) = (1+r)(\hat{n}_1(\theta_1) - \hat{c}_1(\theta_1))$ the net (positive or negative) wealth inherited by agent-2. At date 2, the principal solves the following program $\hat{\mathbf{P}}_\mathbf{2}$:

$$
\begin{aligned}
\hat{\mathbf{P}}_\mathbf{2}: \quad \max_{\{(c_2(\theta_2), n_2(\theta_2))\}} \quad & S_2 = \int_{\underline{\theta}}^{\overline{\theta}} \theta_2\, u(c_2(\theta_2)) - n_2(\theta_2)\, dF(\theta_2) \\
\text{s.t.} \quad & \alpha\, \theta_2\, u(c_2(\theta_2)) - n_2(\theta_2) \geqslant \alpha\, \theta_2\, u(c_2(\tilde{\theta}_2)) - n_2(\tilde{\theta}_2) \quad \forall\ \theta_2,\, \tilde{\theta}_2 \qquad (\hat{\text{IC}}) \\
& c_2(\theta_2) \geqslant 0, \quad n_2(\theta_2) \in [0, \bar{n}] \qquad (\text{F}_2) \\
& c_2(\theta_2) \leqslant n_2(\theta_2) + k(\theta_1) \qquad (\hat{\text{BB}})
\end{aligned}
$$

where $\alpha - 1$ captures the intensity of conflict between motivation and true preferences.

As a benchmark, suppose first that no constrains were imposed on agent-2 (except, naturally, for budget balance). Given a valuation θ_2 and a salience α, agent-2 would

choose the pair of consumption and labor $(x(\theta_2), y(\theta_2))$ that solves:

$$\begin{array}{ll} \max\limits_{\{c_2,n_2\}} & \alpha\,\theta_2\,u(c_2) - n_2 \\ \text{s.t.} & c_2 \geqslant 0, \quad n_2 \in [0,\bar{n}], \quad c_2 \leqslant n_2 + k(\theta_1) \end{array}$$

Assuming an interior solution ($\bar{n}$ sufficiently large) consumption and labor would then be:

$$u'(x(\theta_2)) = \frac{1}{\alpha\,\theta_2} \quad \text{and} \quad y(\theta_2) = x(\theta_2) - k(\theta_1)$$

where $x(\theta_2) > c_2^o(\theta_2)$ for all θ_2. Also, denote $\hat{\theta}$ the cutoff valuation that satisfies:

$$\frac{1}{\hat{\theta}} \times \frac{1 - F(\hat{\theta})}{f(\hat{\theta})} = \frac{\alpha - 1}{\alpha} \tag{C}$$

The principal anticipates agent-2's desired consumption and labor $(x(\theta_2), y(\theta_2))$, which does not coincide with her first-best choice $(c_2^o(\theta_2), n_2^o(\theta_2))$. The solution $(\hat{c}_2(\theta_2), \hat{n}_2(\theta_2))$ to program $\mathbf{\hat{P}_2}$ characterizes the constrained optimum that the cognitive system can achieve at date 2 given the private information and biased motivation of the affective system.

Proposition 4 ***(Asymmetry in the mind with biased motivations at date 2)*** *When agent-2 has private knowledge of his valuation and a biased motivation, the principal constrains only the maximum consumption to $x(\hat{\theta})$ and requires* (B$\hat{\text{B}}$) *to be satisfied.*

Given this restriction, an agent-2 with valuation θ_2 chooses:
- His optimal consumption and labor if $\theta_2 < \hat{\theta}$: $\hat{c}_2(\theta_2) = x(\theta_2)$ and $\hat{n}_2(\theta_2) = y(\theta_2)$
- The optimal consumption and labor of an agent with valuation $\hat{\theta}$ if $\theta_2 \geqslant \hat{\theta}$: $\hat{c}_2(\theta_2) = x(\hat{\theta})$ and $\hat{n}_2(\theta_2) = y(\hat{\theta})$.

Contrary to Proposition 2 where optimal intervention was sophisticated and intrusive, it is now optimal for the principal to follow a simple rule-of-thumb. Since the disagreement is proportional to $(\alpha - 1)\,\theta_2$, the cost of letting agent-2 get away with his desired levels of consumption and labor is small as long as his valuation is low. When the valuation exceeds a certain threshold $\hat{\theta}$, then overconsumption becomes a serious problem and a drastic intervention in the form of a consumption cap becomes optimal. One informal way of interpreting this mechanism is the principal saying "as long as you don't abuse, you can do whatever you want." It is interesting that an extensive use of the 'carrot' (full freedom in the choice of consumption and labor) up to a certain level and then a strict enforcement of the 'stick' (no more choice) above that level endogenously emerges as the best rule.

Note that some nice implications can also be obtained if we apply this problem and mechanism to a completely different setting. Consider for instance a parent (our principal) who can constrain the options available to her offspring (our agent). The offspring privately knows the value he derives from the pleasurable activity, and the parent internalizes only partly his preferences; from her viewpoint, he has a tendency to "party too much" ($\alpha > 1$). Our result suggests that the optimal strategy of the parent is to fully delegate choices to the offspring up to a certain level and firmly intervene above that point (see Corollary 3 below for some comparative statics).

A comment on the optimal mechanism. For the reader familiar with incentive theory, an optimal contract of this form should be somewhat surprising. In fact, we are not aware of any mechanism design problem where, despite the availability of two compensatory tools, the optimal contract is a threshold rule in one variable with full freedom below and bunching above (the techniques to find this optimal mechanism are also somewhat non-standard).[23] The intuition behind this result relies on the fact that incentive compatibility can be satisfied in three different ways. First and trivially, by letting the agent choose his unconstrained optimal pair. Second, by giving every type of agent the same (pooling) contract. Third, by optimally selecting the (monotone) relation between the two variables that induces self-selection. In all problems we know (including the one described in Proposition 2), incentive compatibility of the optimal contract is ensured vía the third criterion or a combination of second and third criteria (when the single crossing property is violated). By contrast, in our setting, there is a tension between incentive compatibility and resource management. On the one hand, self-selection can be induced if the relation between consumption and labor is such that $dn_2/dc_2 = \alpha\theta_2\, u'(c_2)$, see appendix for details. On the other hand, budget balance implies that either $dn_2/dc_2 = 1$ or else some resources are being wasted ($\hat{\text{BB}}$ not binding). Both equalities can simultaneously be satisfied for at most one type. Since the incentives of principal and agent-2 are (not perfectly but still substantially) aligned, it is more important to focus on resource optimization than on self-selection. As a result, incentive compatibility is ensured through the first criterion as long as it is not too costly (θ_2 and/or α small) and then through the second criterion. This also implies that, even in the presence of incentive salience, it is never optimal to

[23] Amador et al. (2004) have a similar cutoff rule although for very different reasons. In their paper, the principal has only one tool (saving target). Thus, contrary to our paper and the traditional mechanism design literature, it is by construction impossible to offer menu of contracts that link two variables (consumption / labor, payment / probability of awarding the good, cost target / transfer, etc.).

discipline the agent by throwing resources away (i.e., by choosing consumption strictly smaller than labor).

Some interesting comparative statics follow from the previous analysis.

Corollary 3 ***(Conflict intensity)***
For a given valuation, agent-2 is less likely to freely choose his desired consumption and labor when the conflict between true and perceived welfare is high and when the willingness to consume is drawn from a less favorable distribution.

As incentive salience or visceral factors become more pronounced, the gap between the principal's first-best choices and agent-2's motivation to work and consume increases, so the former needs to control the latter more tightly in order to avoid excessively inefficient behavior. This is achieved by increasing the range of valuations that are subject to the principal's intervention (formally, $\partial\hat{\theta}/\partial\alpha < 0$). It is worth noting from (C) that $\hat{\theta}(\alpha) < \overline{\theta}$ for all $\alpha > 1$ and $\lim_{\alpha\to 1} \hat{\theta}(\alpha) = \overline{\theta}$: only when the conflict vanishes ($\alpha = 1$) first-best levels can be implemented for all valuations. Stated differently, as soon as true and perceived utility differ (even minimally), the principal is obliged to intervene. Also, for sufficiently strong biases in perceived welfare, it may be optimal to impose a uniform level of consumption and labor ($\hat{\theta} = \underline{\theta}$). Thus, in the parent-offspring interpretation, more intransigent rules just reflect stronger conflicts between the parties involved.

The distribution of valuations also affects intervention. On the one hand, the principal cares about the utility of the agent and proves it by granting full freedom for reasonably small conflicts. On the other hand, she is also concerned about excesses which is why she constrains the agent's choices above a certain level. The optimal rule balances the costs of overconsumption with the costs of pooling. Note that, for a given threshold $\hat{\theta}$, consumption is more likely to be restrictive the more favorable the distribution of valuations (formally, $1 - G(\hat{\theta}) > 1 - F(\hat{\theta})$). In order to avoid an excessive intervention, the principal then becomes more lenient when valuations are more likely to be high.

3.3 Biased motivations at date 1

We can now turn to the analysis of the previous period. From Proposition 4, we know that consumption and labor at date 2 will be adjusted so as to make the budget constraint ($\hat{\text{BB}}$) binding. Although utilizing all resources seems quite natural, we showed that the result is still not obvious, since it comes at the expense of granting excessive freedom (and

therefore excessive consumption) to agent-2.[24] Given this anticipation of future behavior, the program $\hat{\mathbf{P}}_\mathbf{1}$ solved by the principal at date 1 is:

$$\hat{\mathbf{P}}_\mathbf{1}: \max_{\{(c_1(\theta_1),n_1(\theta_1))\}} \quad S_1 = \int_{\underline{\theta}}^{\overline{\theta}} \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) + E_{\theta_2}\Big[\theta_2\, u(\hat{c}_2(\theta_2)) - \hat{n}_2(\theta_2)\Big]\, dF(\theta_1)$$

$$\text{s.t.} \quad \alpha\,\theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) \geqslant \alpha\,\theta_1\, u(c_1(\tilde{\theta}_1)) - n_1(\tilde{\theta}_1) \qquad \forall\ \theta_1,\, \tilde{\theta}_1 \qquad (\hat{\text{IC}})$$

$$c_1(\theta_1) \geqslant 0, \quad n_1(\theta_1) \in [0, \bar{n}] \qquad (\text{F}_1)$$

We can then characterize $(\hat{c}_1(\theta_1), \hat{n}_1(\theta_1))$, the optimal consumption and labor at date 1 from the principal's perspective given private information and salient incentives by agent-1.

Proposition 5 ***(Asymmetry in the mind with biased motivations at date 1)*** *When agent-1 has private knowledge of his valuation and a biased motivation, the principal offers the following menu* $\{(\hat{c}_1(\theta_1), \hat{n}_1(\theta_1))\}_{\theta_1=\underline{\theta}}^{\overline{\theta}}$ *of consumption and labor pairs:*

$$u'(\hat{c}_1(\theta_1)) = \frac{1+r}{\theta_1 + r\,\alpha\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}$$

$$\hat{n}_1(\theta_1) = \bar{n} - \alpha\left[\overline{\theta}\, u(\hat{c}_1(\overline{\theta})) - \theta_1\, u(\hat{c}_1(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} u(\hat{c}_1(x))dx\right]$$

Agent-1 with valuation θ_1 *selects the pair* $(\hat{c}_1(\theta_1), \hat{n}_1(\theta_1))$ *designed for him.*

As we can immediately notice, the options offered by the principal to agent-1 are qualitatively very similar with biased and unbiased motivations ($(\hat{c}_1, \hat{n}_1)$ and (c_1^*, n_1^*) respectively). In both cases, and as in standard mechanism design games, the principal induces self-selection: the higher the valuation of agent-1, the higher will be his choice of both the pleasant activity (consumption) and the unpleasant activity (labor). The reasons for the optimality of this type of mechanism are also the same as described in section 2.2.

There are two implications of this result that are worth being mentioned. First, denote by $c_1^\alpha(\theta_1)$ the consumption by agent-1 under asymmetric information if the principal maximized $V_1 + V_2$, that is, if she shared the biased motivation of the agents. It is straightforward to note that this consumption would solve:[25]

$$u'(c_1^\alpha(\theta_1)) = \frac{1+r}{\alpha\left[\theta_1 + r\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)\right]}$$

[24] This contrasts with Proposition 2 where, given the absence of a conflict at date 2 between the principal and agent-2, it was obvious that (BB) had to be binding.

[25] In fact, it is the consumption determined in Proposition 2 where we replace the function $u(c)$ by $\alpha\, u(c)$.

and therefore that $c_1^{\alpha}(\theta_1) > \hat{c}_1(\theta_1) > c_1^*(\theta_1)$ for all θ_1. In order to induce an agent-1 with an incentive salience to reveal his valuation, the principal must grant him some extra consumption. However, this increase in consumption will not be as high as if the salient motivation was 'real' rather than just 'perceived'. In other words, the forward looking system will neither ignore nor fully integrate the misperceptions of the myopic systems. Instead, it will find the right balance between accommodating biased perceptions and pursuing true motivations.[26]

A second and arguably more interesting conclusion is that the incentive mechanisms proposed at dates 1 and 2 are very different in nature. Consumption at date 1 is always excessively low from agent-1's viewpoint and monotonically increasing in valuation; consumption at date 2 is optimal from agent-2's viewpoint for valuations below a threshold and excessively low and constant for valuations above it. The principal implements different rules because suboptimal choices at dates 1 and 2 have different costs. To be precise, from the principal's viewpoint excessive consumption and insufficient labor at date 1 is not only inefficient per se, it also implies that either fewer resources are left for period 2, or more work is required at period 2, or both. To minimize this cost, a standard "work more if you want to consume more" rule is imposed. By contrast, the concern about the effect of current behavior on future resources disappears at date 2. Excessive consumption by agent-2 is relatively less costly and allowed as long as the agent himself is willing to provide the necessary amount of labor in exchange.

Before concluding, a last clarification is in order. On the one hand, during the entire paper we have been advocating a literal interpretation of our dual-system model: the brain is and must be modelled as a multi-agent organization, with different systems competing to impose their different goals. On the other hand, the revelation games and the results presented in Propositions 2, 4, 5 as well as the optimization processes behind each system should not be taken literally. Just like in most economic optimization problems, they are "as if" mechanisms. An advantage of the normative approach to mechanism design followed in this paper is that it identifies some general properties that can be compared with the behavior observed in practice. In our view, one of the most interesting features of our model is that the qualitative aspects of the optimal mechanisms described in the different propositions have compelling practical interpretations.

[26] Note that as $\alpha \to 1$, choices converge to those under no bias ($\hat{c}_t(\theta_t) \to c_t^*(\theta_t)$ and $\hat{n}_t(\theta_t) \to n_t^*(\theta_t)$).

4 Summary and discussion

The Theory of Organizations has a long tradition of modelling firms as a nexus of agents with incentive problems, informational asymmetries, restricted channels of communication, etc. The main contribution of this paper is to argue, based on neuroscience research, that the brain is also an organization where the different systems play the role of agents. We claim that individual decision-making should be studied from that same multi-agent organization perspective and propose a step in that direction. A few papers have implicitly followed a similar approach before us (Benhabib and Bisin (2004), Bernheim and Rangel (2004)). A main difference is that this literature focuses on automatic processes vs. cognitive optimization whereas we exploit a different neuro-mechanism: the cognitive inaccessibility to our motivations. Starting from this premise, the paper derives a number of results: optimality of self-disciplining rules such as "work more today if you want to consume more today" or "do what you want but don't abuse", foundation for discounting, implications for consumption over the life-cycle, etc.

Our model can be extended in several dimensions. First, we can increase the number of periods. Our conjecture is that every agent will then be less likely to benefit from the existence of private information than its predecessor. It may also generate other testable predictions about discount rates and, in particular, shed light on dynamic preference reversals, as discussed by Fudenberg and Levine (2004) and Loewenstein and O'Donoghue (2004) in related contexts. Second, we can introduce correlated types. This will affect the value of information, and therefore the mechanism offered by the principal. Third, we may attenuate the conflict by assuming that the principal puts a decreasing weight on distant payoffs and that agents have a positive concern for future returns (formally, $1 > \delta_{\text{princ}} > \delta_{\text{agent}} > 0$). Note that the combination of these second and third extensions adds an interesting dimension to the problem: agent-1 will have an incentive to signal his valuation with his consumption and labor choices. It will therefore provide a complementary motive for self-signaling to Bodner and Prelec (2003) and Bénabou and Tirole (2004). Fourth, we can impose strict limitations on the amount of per-period labor $\bar{n}$. This will make full discrimination technically infeasible (the same pair of consumption and labor will have to be offered to a subset of agents with different valuations, what is called "bunching").[27] Last, we may argue that agents can invest resources that increase their productivity of labor. Technically, this will add a moral hazard stage before the contract under asymmetric

[27] Pooling will then occur for reasons that are different than in Proposition 4.

information.

Another, and maybe more promising, alley of research is to test empirically or experimentally some behavioral implications of our theory: the effect of the source of income and the period-to-period labor opportunities on the distribution of consumption over the life cycle, the effect of the distribution of valuations on consumption, etc.

More generally, we have heavily relied on neuroscience experiments to shape our model. However, the lack of evidence (for and against) some ingredients of the model forced us to make debatable choices. For example, is the relation between the cortical and limbic systems best captured by a vertical hierarchy (as we assume) or should we rather think of their interaction as a bargaining process?[28] What is the mechanism behind the principal's ability both to restrict the choices of agents and to keep her end of the deals? Also, there is consistent indication of compartmentalized information within the brain. However, are the systems with restricted access to information aware of these asymmetries? Is private information possessed by the myopic decision-maker? We have assumed that the forward looking system weights equally all periods. Yet, isn't it possible that it also discounts the future (exponentially as in the traditional literature, with a present-bias as the recent evidence suggests or even with a future-bias to compensate for the preferences of the myopic systems)? Last, the existing evidence of incentive salience in humans is suggestive but weak and the evidence in animals provides a poor support for human behavior. Is it possible to obtain stronger support for the biased motivation hypothesis in humans?

Some readers may argue that modelling the brain is excessively bold, at least for the time being, even more so given the concerns and unanswered questions raised above. On the contrary, we believe it is the right time to combine strengths across disciplines and improve our understanding of this complex but fascinating organization called the brain. On one end, multiplying experiments in neuroscience will provide invaluable information to theorists about how to build more accurate organizational models of the brain. On the other end, developing new theories and brain models will help experimental scientists

[28] Bénabou and Pycia (2002) and McClure et al. (2004) advocate bilateral bargaining. Note however that bargaining models are usually sensitive to (i) relative patience of players and (ii) payoffs under no-agreement, two features that raise modelling concerns in our setting. The first one because, by construction, agent-1 is infinitely less patient than the principal and their relation must last exactly one period. The second one, because it is hard to determine each party's "reservation value" or payoff under disagreement (this is precisely why we did not include individual rationality constraints in our model). These among some other considerations led us to opt for the principal-agent approach. Ideally, however, both alternatives should be explored, tested and compared.

determine which hypothesis and assumptions should receive testing priority. Although it is still too early for an assessment, this methodology may eventually lead to a new approach to human decision-making, and we hope that our paper will modestly contribute to stimulate this line of research.

Appendix

Proof of Proposition 2. S_1 can be rewritten as a function of c_1, n_1 and c_2^*:

$$S_1 = \int_{\underline{\theta}}^{\overline{\theta}} \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) + E_{\theta_2}\Big[\theta_2\, u(c_2^*(\theta_2)) - c_2^*(\theta_2) + (1+r)(n_1(\theta_1) - c_1(\theta_1))\Big]\, dF(\theta_1)$$

Given $U_1 = \theta_1\, u(c_1) - n_1$, we can rewrite $\mathbf{P_1^*}$ as a function of (c_1, U_1) instead of (c_1, n_1):

$$\mathbf{P_1^*}: \max_{\{(c_1(\theta_1), U_1(\theta_1))\}} \int_{\underline{\theta}}^{\overline{\theta}} -rU_1(\theta_1) + (1+r)(\theta_1 u(c_1(\theta_1)) - c_1(\theta_1)) + E_{\theta_2}\Big[\theta_2 u(c_2^*(\theta_2)) - c_2^*(\theta_2)\Big] dF(\theta_1)$$

s.t. (IC*) and (F$_1$)

Using standard techniques of implementation theory (see e.g. Fudenberg and Tirole (1991, chapter 7)), we can reduce the continuum of incentive compatibility constraints to a first-order and a second-order condition, namely:

$$\frac{dU_1(\theta_1)}{d\theta_1} = u(c_1(\theta_1)) \tag{a}$$

$$c_1'(\theta_1)\frac{\partial U_1}{\partial n_1}\left[\frac{\partial}{\partial \theta_1}\left(\frac{\partial U_1/\partial c_1}{\partial U_1/\partial n_1}\right)\right] \geqslant 0 \quad \Rightarrow \quad c_1'(\theta_1) \geqslant 0 \tag{b}$$

The key to solve problem $\mathbf{P_1^*}$ is to determine where (IC*) is binding. Since there is no individual rationality constraint, we cannot apply the usual techniques. Given that $n_1(\theta_1) \in [0, \bar{n}]$ for all θ_1, agent-1's utility cannot be smaller than $\underline{U}_1(\theta_1) = \theta_1\, u(c_1(\theta_1)) - \bar{n}$ or greater than $\overline{U}_1(\theta_1) = \theta_1\, u(c_1(\theta_1)) - 0$ for all θ_1. Note that:

$$\frac{d\overline{U}_1}{d\theta_1} = \frac{d\underline{U}_1}{d\theta_1} = u(c_1(\theta_1)) + \theta_1\, u'(c_1(\theta_1)\frac{dc_1}{d\theta_1} > \frac{dU_1}{d\theta_1} = u(c_1(\theta_1)) > 0 \tag{1}$$

This means that for all θ_1, $U_1(\theta_1) \in [\underline{U}_1(\theta_1), \overline{U}_1(\theta_1)]$ and the slope of (IC*) is always smaller than the slopes of $\underline{U}_1(\theta_1)$ and $\overline{U}_1(\theta_1)$. Since, from $\mathbf{P_1^*}$, we know that $U_1(\theta_1)$ enters negatively the principal's objective function, we then have that (IC*) binds at the top or, formally, $U_1(\overline{\theta}) = \underline{U}_1(\overline{\theta})$ which in turn implies that $n_1(\overline{\theta}) = \bar{n}$. Assume that (IC*) does not bind at any other point. Given (1), this is true if $U_1(\underline{\theta}) < \overline{U}_1(\underline{\theta})$ or, equivalently, if $n_1(\underline{\theta}) > 0$. We will neglect this inequality and check it ex-post.

Using (a), and knowing that (IC*) binds at the top, we have:

$$\int_{\theta_1}^{\overline{\theta}} U_1'(x)dx = \int_{\theta_1}^{\overline{\theta}} u(c_1(x))dx \;\Rightarrow\; U_1(\theta_1) = U_1(\overline{\theta}) - \int_{\theta_1}^{\overline{\theta}} u(c_1(x))dx \tag{2}$$

Integrating by parts, we use (a) and (2) to rewrite agent-1's expected utility as:

$$\int_{\underline{\theta}}^{\overline{\theta}} U_1(\theta_1) f(\theta_1) d\theta_1 = U_1(\overline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{F(\theta_1)}{f(\theta_1)}\, u(c_1(\theta_1)) f(\theta_1) d\theta_1$$

Hence, replacing (IC*) by (a) and (b) and introducing (a) in the objective function, we can rewrite $\mathbf{P_1^*}$ as:

$$\max_{c_1(\theta_1)} \int_{\underline{\theta}}^{\overline{\theta}} -r \left[U_1(\overline{\theta}) - \frac{F(\theta_1)}{f(\theta_1)} u(c_1(\theta_1)) \right] + (1+r) \Big[\theta_1 u(c_1(\theta_1)) - c_1(\theta_1) \Big] + A(\theta_2) dF(\theta_1)$$

$$\text{s.t.} \quad c_1'(\theta_1) \geqslant 0, \quad c_1(\theta_1) \geqslant 0, \quad n_1(\theta_1) \in [0, \bar{n}]$$

where $A(\theta_2) = E_{\theta_2}\Big[\theta_2 u(c_2^*(\theta_2)) - c_2^*(\theta_2)\Big]$. First-order condition with respect to c_1 implies:

$$u'(c_1^*(\theta_1)) = \frac{1+r}{\theta_1 + r\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}$$

Note that $dc_1^*(\theta_1)/d\theta_1 > 0$ and $c_1(\theta_1) > 0$ are always satisfied if $\underline{\theta} > 0$. From (2), the second-best level of labor specified in the contract is given by:

$$\theta_1\, u(c_1^*(\theta_1)) - n_1^*(\theta_1) = \overline{\theta}\, u(c_1^*(\overline{\theta})) - \bar{n} - \int_{\theta_1}^{\overline{\theta}} u(c_1(x)) dx$$

that is:

$$n_1^*(\theta_1) = \bar{n} - \left[\overline{\theta}\, u(c_1^*(\overline{\theta})) - \theta_1\, u(c_1^*(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} u(c_1^*(x)) dx \right]$$

and therefore

$$\frac{dn_1^*}{d\theta_1} = \theta_1\, u'(c_1^*(\theta_1))\, \frac{dc_1^*}{d\theta_1} \quad (>0)$$

Last, the neglected inequality $U_1(\underline{\theta}) < \overline{U}_1(\underline{\theta})$ which, in our case, translates into $n_1^*(\underline{\theta}) > 0$ is automatically satisfied if $\bar{n}$ is "sufficiently large" or, more specifically, if:

$$\bar{n} > \overline{\theta}\, u(c_1^*(\overline{\theta})) - \underline{\theta}\, u(c_1^*(\underline{\theta})) - \int_{\underline{\theta}}^{\overline{\theta}} u(c_1^*(x)) dx \;\; = \int_{\underline{\theta}}^{\overline{\theta}} x\, [c_1^*(x)]'\, u'(c_1^*(x)) dx \quad (>0)$$

Proof of Proposition 3. Part (i) follows directly from inspection of $(c_t^*(\theta_t), n_t^*(\theta_t))$ and $(c_t^o(\theta_t), n_t^o(\theta_t))$ in Propositions 1 and 2.

Part (ii). Assuming that $u(c) = c^\rho/\rho$ with $\rho \in (0,1)$, we have:

$$c_1^*(\theta_1) = \left(\frac{\theta_1 + r\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}{1+r} \right)^{\frac{1}{1-\rho}} \quad \text{and} \quad c_1^o(\theta_1) = \left(\frac{\theta_1}{1+r} \right)^{\frac{1}{1-\rho}}$$

and the property $\frac{dc_1^*}{d\theta_1} > \frac{dc_1^o}{d\theta_1} > 0$ follows. The properties $\frac{dn_1^*}{d\theta_1} > 0$ and $\frac{dn_1^o}{d\theta_1} = 0$ also follow directly from Propositions 1 and 2.

From Proposition 2, we know that $\frac{dU_1^*}{d\theta_1} = u(c_1^*(\theta_1)) = \frac{1}{\rho}\left(\frac{\theta_1 + r\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}{1+r}\right)^{\frac{\rho}{1-\rho}}$. Also, $\frac{dU_1^o}{d\theta_1} = u(c_1^o(\theta_1)) + \theta_1 u'(c_1^o(\theta_1))\frac{dc_1^o}{d\theta_1} = \frac{1}{\rho(1-\rho)}\left(\frac{\theta_1}{1+r}\right)^{\frac{\rho}{1-\rho}}$. Therefore

$$\frac{dU_1^*}{d\theta_1} - \frac{dU_1^o}{d\theta_1} \propto (\theta_1 + r(\theta_1 + \tfrac{F(\theta_1)}{f(\theta_1)}))^{\frac{\rho}{1-\rho}} - \tfrac{1}{1-\rho}(\theta_1)^{\frac{\rho}{1-\rho}}$$

and the last inequality as a function of r follows.

Part (iii). $c_1^*(\theta_1) > c_1^{**}(\theta_1)$ is immediate given $\frac{F(\theta_1)}{f(\theta_1)} > \frac{G(\theta_1)}{g(\theta_1)}$. Also, from (IC*):

$$\left.\frac{\partial U_1(\tilde{\theta}_1, \theta_1)}{\partial \tilde{\theta}_1}\right|_{\tilde{\theta}_1 = \theta_1} = 0 \quad \Rightarrow \quad \theta_1\, u'(c_1(\theta_1)) c_1'(\theta_1) = n_1'(\theta_1)$$

Therefore, $\left.\frac{dn_1^{**}}{dc_1^{**}}\right|_{\theta_1} = \theta_1\, u'(c_1^{**}) > \left.\frac{dn_1^*}{dc_1^*}\right|_{\theta_1} = \theta_1\, u'(c_1^*) > 0$.

Proof of Proposition 4. Using the same methodology as in Proposition 2, we can rewrite the (IĈ) constraints of $\hat{\mathbf{P}}_\mathbf{2}$ as a first-order and a second-order condition:

$$V_2'(\theta_2) = \alpha\, u(c_2(\theta_2)) \qquad \text{and} \qquad c_2'(\theta_2) \geqslant 0$$

where, by abuse of notation, we have denoted $V_2'(\cdot)$ the total differential of V_2 with respect to θ_2. Define $H(\theta_2) = \alpha\,\theta_2\, u(c_2(\theta_2)) - \bar{n}$ and $I(\theta_2) = \alpha\,\theta_2\, u(c_2(\theta_2)) - c_2(\theta_2) + k(\theta_1)$. Ignoring for the moment the constraint $n_2(\theta_2) \geqslant 0$, we can rewrite program $\hat{\mathbf{P}}_\mathbf{2}$ as:

$$\begin{aligned}
\hat{\mathbf{P}}_\mathbf{2}: \max_{\{(c_2(\theta_2), V_2(\theta_2))\}} \quad & S_2 = \int_{\underline{\theta}}^{\overline{\theta}} V_2(\theta_2) - (\alpha - 1)\theta_2\, u(c_2(\theta_2))\, dF(\theta_2) & \\
\text{s.t.} \quad & V_2'(\theta_2) = \alpha\, u(c_2(\theta_2)) & (\hat{\text{IC}}_1)\\
& c_2'(\theta_2) \geqslant 0 & (\hat{\text{IC}}_2)\\
& c_2(\theta_2) \geqslant 0 & (\text{F}_\text{c})\\
& V_2(\theta_2) \geqslant H(\theta_2) & (\text{F}_\text{n})\\
& V_2(\theta_2) \leqslant I(\theta_2) & (\hat{\text{BB}})
\end{aligned}$$

where we have rearranged the objective function so that $V_2(\cdot)$ appears on it, split (IĈ) into a first and a second order condition (IĈ$_1$) and (IĈ$_2$), rewritten the feasibility constraint

$n_2(\theta_2) \leqslant \bar{n}$ in terms of $V_2(\cdot)$ and $H(\cdot)$ and called it ($\mathrm{F_n}$), and rewritten the budget balance constraint in terms of $V_2(\cdot)$ and $I(\cdot)$. Note that if ($\hat{\mathrm{IC}}_2$) is satisfied, then:

$$H'(\theta_2) = \alpha\, u(c_2(\theta_2)) + \alpha\, \theta_2\, u'(c_2(\theta_2))\, c_2'(\theta_2) \geqslant V_2'(\theta_2)$$

We also have

$$I'(\theta_2) = \alpha\, u(c_2(\theta_2)) + \Big[\alpha\, \theta_2\, u'(c_2(\theta_2)) - 1\Big] c_2'(\theta_2)$$

Recall that $x(\theta_2)$ is the consumption that solves agent-2's unconstrained optimization problem. Formally, $\alpha\, \theta_2\, u'(x(\theta_2)) - 1 = 0$. Assuming that ($\hat{\mathrm{IC}}_2$) is satisfied, then:

- $I'(\theta_2) > V_2'(\theta_2)$ if $c_2'(\theta_2) > 0$ and $c_2(\theta_2) < x(\theta_2)$.
- $I'(\theta_2) = V_2'(\theta_2)$ if $c_2'(\theta_2) = 0$ or $c_2(\theta_2) = x(\theta_2)$.
- $I'(\theta_2) < V_2'(\theta_2)$ if $c_2'(\theta_2) > 0$ and $c_2(\theta_2) > x(\theta_2)$.

Suppose that, in equilibrium, $c_2(\theta_2) > x(\theta_2)$. This means that $n_2(\theta_2) > x(\theta_2) - k(\theta_1)$. Even if it were incentive-compatible, this solution would yield lower utility both to the principal and to agent-2 than $c_2(\theta_2) = x(\theta_2)$ and $n_2(\theta_2) = x(\theta_2) - k(\theta_1)$. So, in equilibrium, we must have $c_2(\theta_2) \leqslant x(\theta_2)$ and therefore $I'(\theta_2) \geqslant V_2'(\theta_2)$.

Technically, the key difference between $\hat{\mathbf{P}}_\mathbf{2}$ and a standard mechanism design problem is that $V_2(\theta_2)$ enters positively the objective function. Therefore, the principal *maximizes* agent-2's rents subject to ($\mathrm{F_n}$) and ($\hat{\mathrm{BB}}$), that is subject to $V_2(\theta_2) \in [H(\theta_2), I(\theta_2)]$. Given that $H'(\theta_2) > V_2'(\theta_2)$ and $I'(\theta_2) \geqslant V_2'(\theta_2)$, then $V_2(\theta_2)$ binds at the bottom on $I(\theta_2)$, so

$$n_2(\underline{\theta}) = c_2(\underline{\theta}) - k(\theta_1) \quad \text{and} \quad V_2(\underline{\theta}) = \alpha\, \underline{\theta}\, u(c_2(\underline{\theta})) - c_2(\underline{\theta}) + k(\theta_1)$$

Using ($\hat{\mathrm{IC}}_1$) and with $V_2(\theta_2)$ binding at $\theta_2 = \underline{\theta}$, the function can be rewritten as:

$$V_2(\theta_2) = V_2(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_2} \alpha\, u(c_2(s))ds \tag{3}$$

Standard integration by parts implies that agent-2's expected utility is:

$$\int_{\underline{\theta}}^{\overline{\theta}} V_2(\theta_2) f(\theta_2) d\theta_2 = V_2(\underline{\theta}) + \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta_2)}{f(\theta_2)}\, \alpha\, u(c_2(\theta_2)) f(\theta_2) d\theta_2$$

so program $\hat{\mathbf{P}}_\mathbf{2}$ becomes:

$$\begin{aligned} \hat{\mathbf{P}}_\mathbf{2}: \max_{c_2(\theta_2)} \quad & S_2 = \int_{\underline{\theta}}^{\overline{\theta}} \frac{1 - F(\theta_2)}{f(\theta_2)}\, \alpha\, u(c_2(\theta_2)) - (\alpha - 1)\theta_2\, u(c_2(\theta_2))\, dF(\theta_2) + V_2(\underline{\theta}) \\ \text{s.t.} \quad & c_2'(\theta_2) \gtrless 0, \quad c_2(\theta_2) \leqslant x(\theta_2), \quad c_2(\theta_2) \geqslant 0 \end{aligned}$$

the derivative of the objective function with respect to $c_2(\theta_2)$ is:

$$\left[\alpha\frac{1-F(\theta_2)}{f(\theta_2)}-(\alpha-1)\theta_2\right]u'(c_2(\theta_2))$$

Denote $Z(\theta_2,\alpha)=\alpha\frac{1-F(\theta_2)}{f(\theta_2)}-(\alpha-1)\theta_2$. Since $\frac{d}{d\theta}\left[\frac{1-F(\theta)}{f(\theta)}\right]<0$, then $\frac{\partial Z(\theta_2,\alpha)}{\partial\theta_2}<0$, $Z(\overline{\theta},1)=0$ and $Z(\overline{\theta},\alpha)<0$ for all $\alpha>1$. Hence, for all $\alpha>1$, there exists $\hat{\theta}(\alpha)\in[\underline{\theta},\overline{\theta})$ such that $Z(\hat{\theta}(\alpha),\alpha)=0$, $Z(\theta_2,\alpha)>0$ for all $\theta_2<\hat{\theta}(\alpha)$ and $Z(\theta_2,\alpha)<0$ for all $\theta_2>\hat{\theta}(\alpha)$.[29]

Now, fix α (>1). Using the properties of the function $Z(\cdot)$, we get:

(i) For all $\theta_2<\hat{\theta}$, the objective function is strictly increasing in $c_2(\theta_2)$, therefore optimal consumption is $\hat{c}_2(\theta_2)=x(\theta_2)$. Using (3), optimal labor is then:

$$\hat{n}_2(\theta_2)=\alpha\,\theta_2\,u(x(\theta_2))-\int_{\underline{\theta}}^{\theta_2}\alpha\,u(x(s))ds-\alpha\,\underline{\theta}\,u(x(\underline{\theta}))+x(\underline{\theta})-k(\theta_1) \tag{4}$$

Note that agent-2's equilibrium utility at his preferred solution $(x(\theta_2),x(\theta_2)-k(\theta_1))$ is:

$$W_2(\theta_2)=\alpha\,\theta_2\,u(x(\theta_2))-x(\theta_2)+k(\theta_1)\qquad\text{with}\quad W_2'(\theta_2)=\alpha\,u(x(\theta_2))$$

which can be rewritten as:

$$W_2(\theta_2)=\int_{\underline{\theta}}^{\theta_2}\alpha\,u(x(s))ds+W_2(\underline{\theta})\qquad\text{where}\quad W_2(\underline{\theta})=\alpha\,\underline{\theta}\,u(x(\underline{\theta}))-x(\underline{\theta})+k(\theta_1) \tag{5}$$

Combining (4) and (5) we finally obtain:

$$\hat{n}_2(\theta_2)=\alpha\,\theta_2\,u(x(\theta_2))-W_2(\theta_2)\quad\Rightarrow\quad\hat{n}_2(\theta_2)=x(\theta_2)-k(\theta_1)$$

(ii) For all $\theta_2>\hat{\theta}$, the objective function is strictly decreasing in $c_2(\theta_2)$, so the principal chooses the smallest possible consumption that satisfies $c_2'(\theta_2)\geqslant 0$. In the optimum then $\hat{c}_2'(\theta_2)=0$ and $\hat{c}_2(\theta_2)=x(\hat{\theta})$. Also,

$$\hat{n}_2(\theta_2)=\alpha\,\theta_2\,u(x(\hat{\theta}))-\int_{\underline{\theta}}^{\hat{\theta}}\alpha\,u(x(s))ds-\int_{\hat{\theta}}^{\theta_2}\alpha\,u(x(\hat{\theta}))ds-\alpha\,\underline{\theta}\,u(x(\underline{\theta}))+x(\underline{\theta})-k(\theta_1)$$

Using the same techniques as before, we finally obtain:

$$\hat{n}_2(\theta_2)=x(\hat{\theta})-k(\theta_1)$$

[29] Note that there is a mathematical abuse in this formulation: if $\underline{\theta}f(\underline{\theta})>1$ and $\alpha>\bar{\alpha}\equiv\frac{\underline{\theta}f(\underline{\theta})}{\underline{\theta}f(\underline{\theta})-1}$ then $Z(\theta_2,\alpha)<0$ for all $\theta_2\in[\underline{\theta},\overline{\theta}]$.

Two final remarks. First, $\hat{n}_2'(\theta_2) \geqslant 0$, so $\hat{n}_2(\theta_2) \geqslant 0$ reduces to $\hat{n}_2(\underline{\theta}) \geqslant 0$. For a suitably chosen $\bar{n}$, this constraint is automatically satisfied (e.g., a sufficient condition is $k(\theta_1) < 0$ for all θ_1). Second, given that $(x(\theta_2), x(\theta_2) - k(\theta_1))$ is optimal for agent-2, it is obvious that the mechanism $(\hat{c}_2(\theta_2), \hat{n}_2(\theta_2))$ can be implemented just by constraining maximum consumption to $x(\hat{\theta})$ and imposing $n_2(\theta_2) \geqslant c_2(\theta_2) - k(\theta_1)$.

Proof of Corollary 3. $Z(\hat{\theta}(\alpha), \alpha) = 0$ can be rewritten as in equation (C). Since the left hand side is decreasing in $\hat{\theta}$ and the right hand side is increasing in α, then $\frac{\partial \hat{\theta}}{\partial \alpha} < 0$. Also, recall from footnote 29 that there exists $\bar{\alpha}$ such that if $\underline{\theta} f(\underline{\theta}) > 1$ and $\alpha > \bar{\alpha}$ then $Z(\underline{\theta}, \alpha) < 0$, in which case $\hat{\theta} = \underline{\theta}$ and pooling occurs for all types.

Denote by $\check{\theta}$ the valuation that solves $Z(\check{\theta}, \alpha) = 0$ when θ_2 is drawn from distribution $G(\cdot)$ rather than $F(\cdot)$. Therefore, $\hat{\theta}$ and $\check{\theta}$ solve respectively $\alpha \frac{1-F(\hat{\theta})}{f(\hat{\theta})} = (\alpha - 1)\hat{\theta}$ and $\alpha \frac{1-G(\check{\theta})}{g(\check{\theta})} = (\alpha - 1)\check{\theta}$. If $\left(\frac{g(\theta_t)}{f(\theta_t)}\right)' > 0$ then $\frac{g(\theta)}{1-G(\theta)} < \frac{f(\theta)}{1-F(\theta)}$ (see footnote 17), which necessarily implies that $\check{\theta} > \hat{\theta}$.

Proof of Proposition 5. It follows exactly the same techniques as the proof of Proposition 2, so we will skip most steps. $\hat{\mathbf{P}}_\mathbf{1}$ can be rewritten as a function of (c_1, V_1):

$$\hat{\mathbf{P}}_\mathbf{1}: \max_{\{(c_1(\theta_1), V_1(\theta_1))\}} \int_{\underline{\theta}}^{\overline{\theta}} -rV_1(\theta_1) + (1 + r\alpha)\,\theta_1\,u(c_1(\theta_1)) - (1+r)c_1(\theta_1) + B(\theta_2)\,dF(\theta_1)$$

$$\text{s.t.} \quad \frac{dV_1(\theta_1)}{d\theta_1} = \alpha\,u(c_1(\theta_1)), \quad c_1'(\theta_1) \geqslant 0, \quad c_1(\theta_1) \geqslant 0, \quad n_1(\theta_1) \in [0, \bar{n}]$$

where $B(\theta_2) = E_{\theta_2}\Big[\theta_2 u(\hat{c}_2(\theta_2)) - \hat{c}_2(\theta_2)\Big]$. Integrating by parts the first-order condition of (ÎC), we have $\int_{\underline{\theta}}^{\overline{\theta}} V_1(\theta_1)dF(\theta_1) = V_1(\overline{\theta}) - \int_{\underline{\theta}}^{\overline{\theta}} \frac{F(\theta_1)}{f(\theta_1)}\,\alpha\,u(c_1(\theta_1))f(\theta_1)d\theta_1$. Incorporating it in the objective function and maximizing with respect to $c_1(\theta_1)$, we get:

$$u'(\hat{c}_1(\theta_1)) = \frac{1+r}{\theta_1 + r\,\alpha\left(\theta_1 + \frac{F(\theta_1)}{f(\theta_1)}\right)}$$

Using $V_1(\theta_1) = V_1(\overline{\theta}) - \int_{\theta_1}^{\overline{\theta}} \alpha\,u(c_1(x))dx$ and assuming $\bar{n} > \alpha \int_{\underline{\theta}}^{\overline{\theta}} x\,[\hat{c}_1(x)]'\,u'(\hat{c}_1(x))dx$, then

$$\hat{n}_1(\theta_1) = \bar{n} - \alpha\left[\overline{\theta}\,u(\hat{c}_1(\overline{\theta})) - \theta_1\,u(\hat{c}_1(\theta_1)) - \int_{\theta_1}^{\overline{\theta}} u(\hat{c}_1(x))dx\right] > 0 \quad \forall\,\theta_1$$

and the result follows.

References

1. Amador, M., Werning, I and G.M. Angeletos (2004), "Commitment vs. Flexibility", *mimeo*, MIT and Stanford.

2. Baumeister, R. (2003), "The Psychology of Irrationality: Why People Make Foolish, Self-Defeating Choices", in I. Brocas and J. Carrillo *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, 3-16, Oxford: Oxford University Press.

3. Bénabou, R. and M. Pycia (2002), "Dynamic Inconsistency and Self-Control: A Planner-Doer Interpretation", *Economics Letters*, 77, 419-424.

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

5. Bénabou, R. and J. Tirole (2004), "Willpower and Personal Rules", *Journal of Political Economy*, 112, 848-887.

6. Benhabib, J. and A. Bisin (2004), "Modelling Internal Commitment Mechanisms and Self-Control: a Neuroeconomics Approach to Consumption-Saving Decisions", forthcoming in *Games and Economic Behavior*.

7. Bernheim, B.D. and A. Rangel (2004), "Addiction and Cue-Triggered Decision Processes", forthcoming in *American Economic Review*.

8. Berns, G, Cohen, D. and M. Mintun (1997), "Brain Regions Responsive to Novelty in the Absence of Awareness", *Science*, 276, 1272-1275.

9. Berridge, K. (2003), "Irrational Pursuit: Hyper-Incentives from a Visceral Brain", in I. Brocas and J. Carrillo eds. *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, 17-40, Oxford: Oxford University Press.

10. Bodner, R. and D. Prelec (2003), "Self-Signaling and Diagnostic Utility in Everyday Decision Making" in I. Brocas and J. Carrillo *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, 105-126, Oxford: Oxford University Press.

11. Botvinick, M., Braver, T., Barch, D., Carter, C., and J. Cohen (2001), "Conflict Monitoring and Cognitive Control", *Psychological Review*, 108(3), 624-652

12. Brocas, I. and J.D. Carrillo (2004), "Entrepreneurial Boldness and Excessive Investment", *Journal of Economics & Management Strategy*, 13, 321-50.

13. Caillaud, B. and B. Jullien (2000), "Modelling Time-Inconsistent Preferences", *European Economic Review*, 44, 1116-1124.

14. Caillaud, B., Cohen, D. and B. Jullien (1999), "Towards a Theory of Self-Restraint", *mimeo*, CERAS and Toulouse.

15. Camerer, C., Babcock, L., Loewenstein, G. and R. Thaler (1997), " Labor Supply of New York City Cabdrivers: One Day at a Time", *Quarterly Journal of Economics*, 112, 407-441.

16. Camerer, C., Loewenstein, G. and D. Prelec (2004a), "Neuroeconomics: How Neuroscience can Inform Economics", forthcoming in *Journal of Economic Literature*.

17. Camerer, C., Loewenstein, G. and D. Prelec (2004b), "Neuroeconomics: Why Economics Needs Brains", forthcoming in *Scandinavian Journal of Economics*.

18. Caplin, A. and J. Leahy (2001), "Psychological Expected Utility Theory and Anticipatory Feelings", *Quarterly Journal of Economics*, 116, 55-80.

19. Caplin, A. and J. Leahy (2004), "The Social Discount Rate", forthcoming in *Journal of Political Economy*.

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

21. Carter, C., Braver, T., Barch, D., Botvinick, M. Noll, D. and J. Cohen (1998), "Anterior Cingulate Cortex, Error Detection, and the Online Monitoring of Performance", *Science*, 280, 747-749.

22. Damasio, A. (1994), *Descartes' Error: Emotion, Reason and the Human Brain*, New York: G.P. Putnam.

23. Elster, J. (2004), "Costs and Constraints in the Economy of the Mind", in I. Brocas and J.D. Carrillo eds. *The Psychology of Economic Decisions. Vol.2: Reasons and Choices*, pp. 3-14, Oxford: Oxford University Press.

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

25. Freeman, W. (1997), "Self, Awareness of Self, and the Illusion of Control", *Behavioral and Brain Sciences*, 20, 112-113.

26. Freud, S. (1927), *The Ego and the Id*, London: Hogarth.

27. Fudenberg, D and D.K. Levine (2004), "A Dual Self Model of Impulse Control", *mimeo*, Harvard and UCLA.

28. Fudenberg, D. and J. Tirole (1991), *Game Theory*, Cambridge: MIT Press.

29. Gul, F. and W. Pesendorfer (2001), "Temptation and Self-Control", *Econometrica*, 69, 1403-1435.

30. Gur, R. and H. Sackeim (1979), "Self-deception: A Concept in Search of a Phenomenon", *Journal of Personality and Social Psychology*, 37, 147169.

31. Hume, D. (1739), *A Treatise of Human Nature*, New York: Oxford University Press.

32. Laffont, J.J. and D. Martimort (1997), "Collusion under Asymmetric Information", *Econometrica*, 65, 875-911.

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

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

35. Loewenstein, G. and T. O'Donoghue (2004), "Animal Spirits: Affective and Deliberative Processes in Economic Behavior", *mimeo*, Carnegie Mellon and Cornell.

36. Loewenstein, G., Weber, W., Flory, J., Manuck, S. and M. Muldoon (2001), "Dimensions of time-discounting", *mimeo*, Carnegie Mellon.

37. McClure, S., Laibson, D., Loewenstein, G. and J. Cohen (2004), "Separate Neural Systems Value Immediate and Delayed Monetary Rewards", *Science*, 306, 503-507.

38. Montague, R., Hyman, S. and J. Cohen (2004), "Computational Roles for Dopamine in Behavioral Control", *Nature*, 4XX, 2-9.

39. Ochsner, K. and J. Gross (2005) "The cognitive control of emotion", *Trends in Cognitive Sciences*, 9(5), 242-249.

40. Palacios-Huerta, I. (2004), "Consistent Intertemporal Decision-Making through Memory and Anticipation", in I. Brocas and J. Carrillo eds. *The Psychology of Economic Decisions. Vol 2: Reasons and Choices*, pp. 67-80, Oxford: Oxford University Press.

41. Rabin, M. (1998), "Psychology and Economics", *Journal of Economic Literature*, 36, 11-46.

42. Pascal, B. (1670), *Les Pensées*.

43. Shefrin, H.M and Thaler, R.H. (1988), "The Behavioral Life-Cycle Hypothesis", *Economic Inquiry*, 26, 609-643.

44. Smith, A. (1759), *The Theory of Moral Sentiments*, London: Millar.

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

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

47. Tirole, J. (2002), "Rational Irrationality: Some Economics of Self-Management", *European Economic Review*, 46, 633-655.

48. Whalen, P., Rauch, S., Etcoff, N., McInerney, S., Lee, M., and M. Jenike (1998), "Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge", *The Journal of Neuroscience*, 18(1), 411-418.