The Brain as a Hierarchical Organization

by **Isabelle Brocas** and **Juan D. Carrillo** *

Abstract

Based on recent neuroscience evidence, we model the brain as a dual-system organization subject to three conflicts: asymmetric information, temporal horizon and incentive salience. Under the first and second conflicts, we show that the uninformed system imposes a positive link between consumption and labor at every period. Furthermore, decreasing impatience endogenously emerges as a consequence of these two conflicts. Under the first and third conflicts, it becomes optimal to set a consumption cap. Finally, we discuss the behavioral implications of these rules for choice bracketing and expense tracking, and for consumption over the life-cycle. (JEL D82, D87).

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

Economics has experienced an inflow of fresh ideas following an addition of elements from psychology into economic models. A recent literature incorporates intrapersonal tensions into these models. The present paper provides a step in this direction. Our basic premise is the existence of three types of brain based conflicts. First, a conflict between the information available in different areas of the brain, which we refer to as an "asymmetric information conflict." Second, a conflict between the importance attached to temporally close versus temporally distant events, which we refer to as a "temporal horizon conflict." And third, a conflict between the relative weight in utility attached to tempting versus non-tempting goods, which we refer to as an "incentive salience conflict." Starting from these three assumptions about the architecture of the brain, we construct an

*Brocas: Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089-0253, and CEPR (e-mail: brocas@usc.edu). Carrillo: Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089-0253 and CEPR (e-mail: juandc@usc.edu). We thank Roland Bénabou, Colin Camerer, Timothy Derdenger, Stefano DellaVigna, Christian Hellwig, Yong Kim, Botond Koszegi, George Loewenstein, John O'Doherty, Ignacio Palacios-Huerta, Drazen Prelec, Matthew Rabin, Antonio Rangel, John Riley, Hersh Shefrin, Bill Zame and seminar participants at USC, Princeton, Columbia, Toulouse, Stanford SITE, Caltech, UC Davis, UC Los Angeles and UC Berkeley for comments and suggestions. We are especially thankful to Antoine Bechara and Paul Glimcher for their guidance and patience.

orthodox multi-period, multi-action model. The model is solved with tools adapted from mechanism design and is used to provide foundations for discounting and an explanation for several behavioral anomalies.

Asymmetric information and temporal conflicts is the focus of section 2. We consider an individual who undertakes two activities during several periods, one pleasant (consumption) and one unpleasant (labor). Activities are linked through an intertemporal budget constraint. To model the temporal and informational conflicts, we divide the individual into an impulsive/myopic system (the agent, he) and a cognitive/forward-looking system (the principal, she). We then assume that the marginal value of consumption varies from period to period and is only known by the agent. Despite the fact that the cognitive system has control over the impulsive system, she cannot impose her first-best choices due to the informational conflict. Instead, she proposes a menu of pairs where the levels of both activities are positively linked within each period, allowing the agent to signal which of these pairs he prefers. Thus, we show that a self-disciplining intrapersonal rule of behavior of the form "work more today if you want to consume more today" emerges endogenously (Propositions 1 and 2). The consumption pattern exhibits properties that are consistent with modern behavioral theories of choice over time: decreasing impatience and different degrees of impatience for different categories of activities (Proposition 3). Thus, discounting is *derived* from the primitives of our model (informational asymmetry) rather than assumed as an intrinsic feature of preferences.

The behavioral implications of the model are discussed in section 3. First, our theory rationalizes narrow choice bracketing, a practice based on local rather than global optimization that standard models have problems explaining (Daniel Read, George Loewenstein and Matthew Rabin 1999). Indeed, by separating consumption into arbitrarily defined categories and imposing a negative relationship between expenditures on each of them, the principal can achieve financial discipline. Second, our psychological personal rule can help understand some empirical findings difficult to reconcile with the theory of intertemporal consumption (Hersh Shefrin and Richard Thaler 1988). In particular, our rule predicts that consumption tracks earned income, simply because self-discipline can be more easily implemented in periods with better access to labor. The rule also predicts an imperfect substitutability between mandatory and discretionary savings.

In section 4, we abstract from the temporal dimension and focus on the informational and incentive conflicts. In this case, the individual must allocate resources between a tempting good and a non-tempting good. The relative desirability of the tempting good

is only known to the agent. We formalize the concept of 'incentive salience' by assuming that the agent has a biased motivation compared to the fundamental preferences of the principal. Namely, the agent is willing to engage in excessive consumption of the tempting good. When the degree of the conflict increases with the desirability of the tempting good, it is optimal for the principal to impose a consumption cap. That is, she sets a non-intrusive rule of the form "do what you want as long as you don't abuse." When the degree of conflict decreases in the desirability of the tempting good, it may become optimal to waste resources as a commitment device against incurring excesses (Propositions 4 and 5).

The main justifications for our informational, temporal and incentive conflicts in the brain come from neuroscientific research. Section 1 reviews this evidence.[1] The existing literature in psychology and, to a lesser extent, economics has also addressed these issues. The remainder of this section summarizes some findings.

Although controversial in economics, informational conflicts within the individual are widely accepted in other disciplines. Some influential theories in social psychology rely on this assumption. Cognitive dissonance (Leon Festinger 1957) is based on the idea that an individual can simultaneously hold two contradictory beliefs. When this happens, the person acts upon one of them to reduce the discomfort created by such inconsistency. According to the theory of self-deception (Ruben Gur and Harold Sackeim 1979), one of these contradictory beliefs may not be subject to awareness, and this unawareness will be motivated. Self-perception theory (Daryl Bem 1967) makes a stronger statement: individuals do not have the capability to observe directly their own attitudes and therefore, they need to infer them from their emotions and other internal states. In other words, the individual is like an outside observer who relies on external cues to learn his inner states. As for economics, Ronit Bodner and Prelec (2003) is the only existing formal study of asymmetric information within the individual. The authors focus on self-signaling, or how the gut who possesses some information that cannot be introspected by the mind uses actions to signal preferences to himself. A different but related idea can be found in the literature on the construction of preferences. Recent experimental evidence suggests that preferences for ordinary products are unknown and malleable, even after sampling (Dan Ariely, Loewenstein and Prelec 2003, 2006). Several theories have been proposed to understand how preferences are constructed over time, through experience, and with

[1]For summaries of how neuroscience can help economics, see Colin Camerer, Loewenstein and Drazen Prelec (2004, 2005).

the help of memory processes (Sarah Lichtenstein and Paul Slovic (2006, part V)). Under this interpretation, our model argues that, in the process of constructing preferences, the impulsive part of the individual should not be repressed. Instead, it should be permitted to make (optimally designed) constrained choices that facilitate the revelation of current preferences while reducing their possible negative effects on future preferences.

Temporal conflicts have also been stressed in psychology (see e.g. George Ainslie 1992). They are somewhat more accepted in economics than informational conflicts, either under hyperbolic discounting (Robert Strotz (1956), David Laibson (1997) and others) or under some other formulation of the self-control problem (Bernard Caillaud, Daniel Cohen and Bruno Jullien (1999), Faruk Gul and Wolfgang Pesendorfer (2001) and others).[2] A strand of this literature has studied the effects of imperfect self-knowledge on decision-making.[3] In these studies, the temporal and informational conflicts occur between periods. Instead, we stress the existence of these conflicts *within* each period. Hence, the view of the brain as a multi-system organization. In this respect, our paper is closer to Thaler and Shefrin (1981) and Shefrin and Thaler (1988), to our knowledge the first studies which divided the individual into two entities, one myopic and one forward-looking. These articles explain the benefits of commitment devices such as mandatory pension plans and lump-sum bonuses in promoting savings. They have been extended and further developed by Drew Fudenberg and David Levine (2006) and Loewenstein and Edward O'Donoghue (2005). The first paper argues that the split-self approach can explain dynamic preference reversals and the paradox of risk-aversion in the large and in the small. The second shows that this framework sets a parsimonious benchmark to study the optimal decision to exert willpower. None of these papers, however, consider asymmetric information or incentive salience, two key driving forces of our analysis.

Finally, the biasing role of affect on cognition has received a growing interest across disciplines. It has been argued that the affective system helps (Antonio Damasio, 1994), constrains (Jon Elster, 2004) or prevents (Roy Baumeister, 2003) the cognitive system from making optimal choices. Loewenstein (1996) argues that emotions and drives cause individuals to behave contrary to their long-term interest. This dichotomy between im-

[2]See Caillaud and Jullien (2000) for a review of different ways to model time-inconsistent preferences, Andrew Caplin and John Leahy (2001) for the time-inconsistency effect generated by anticipatory feelings and Roland Bénabou and Marek Pycia (2002) for a planner-doer reinterpretation of the self control problem.

[3]See e.g. Carrillo and Thomas Mariotti (2000), Brocas and Carrillo (2004), Bénabou and Jean Tirole (2004), Marco Bataglini, Bénabou and Tirole (2005) and Manuel Amador, Ivan Werning and George-Marios Angeletos (2006).

pulsive and reflective behavior has also been the object of neuroeconomic research. Jess Benhabib and Alberto Bisin (2005) study the consumption choice of an individual who can invoke either a costless automatic process which is susceptible to temptation or a costly control process which is immune to temptation. B. Douglas Bernheim and Antonio Rangel (2004) analyze addiction under the assumption that the individual operates in either a 'cold mode' where he selects his preferred alternative or a 'hot mode' where choices may be suboptimal given preferences. Note that, in these dual-system models, information is complete. Impulsive choices are automatic responses to shocks or cues. By contrast, in our model, the agent optimizes according to a well-defined goal, only his motivation is biased. Because of his superior information, the agent may end up affecting choices. In that respect, our static model with incentive salience and two activities is formally closer to the model by Amador, Werning and Angeletos (2006), where the conflict is based on hyperbolic discounting and the two activities are consumption at different dates. Under some conditions, we replicate the main conclusion of that paper, namely the second-best optimality of a consumption cap.

1 Conflicts in the brain: some evidence from neuroscience

Brain modularity is a well-accepted neurobiological fact.[4] There is also ample evidence that brain systems are often in competition and conflict.[5] As discussed above, the basic premises of our analysis are the existence of informational, temporal and incentive conflicts in the brain. We proceed with a brief review of the evidence in neuroscience that supports each of these conflicts as well as the connections among them.

1. Asymmetric Information. Although not heavily emphasized in the neuroeconomics literature, asymmetric information is, for purely anatomical and evolutionary reasons, arguably the least controversial of the conflicts proposed here. Neural connectivity is a strongly limited resource that evolution spends sparingly. As a result, most brain areas are unidirectionally connected to others. These restrictions *physiologically* constrain the

[4]By contrast, it has been demonstrated by anatomists and neuroscientists that, contrary to the popular view based on theories developed in the 1940s and 1950s, reason and emotion *do not* pertain to two distinct brain systems (see Joseph LeDoux (1996, ch. 4) for a non-technical historical perspective).

[5]See for example the reviews by Russell Poldrack and Paul Rodriguez (2004) on competition between memory systems and Earl Miller and Jonathan Cohen (2001) on competition between information processing systems.

flow of information. Neuroscientific research provides many examples of informational asymmetries using brain imaging techniques (PET scan and fMRI). Studies have shown activation of the ventral striatum, right striatum and amygdala in response to novelty, implicit learning and fear, in each case without conscious awareness of subjects (see Gregory Berns, Jonathan Cohen and Mark Mintum (1997), Scott Rauch et al. (1997) and Paul Whalen et al. (1998), respectively). Research on individuals with brain lesions reveals similar dissociations. Despite their having an intact declarative memory, patients with damage in the neostriatum and the amygdala exhibit, respectively, an impaired ability for gradual learning and an impaired capacity to acquire conditioned responses to emotional stimuli (Barbara Knowlton, Jennifer Mangels and Larry Squire (1996), Antoine Bechara et al. (1995)).

2. Temporal horizon. The evidence of a time-evaluation conflict is more indirect, and yet more popular in neuroeconomics. On the far-sighted end, Damasio (1994) demonstrates that damage in the ventromedial prefrontal cortex impairs the ability of patients to engage in long term planning. This severe myopia is confirmed by Bechara et al. (1999) using a gambling task experiment. On the short-sighted end, LeDoux (1996) shows that the amygdala plays a crucial role in the expression of impulsive, emotional behavior. Bechara et al. (1999) conclude that patients with lesions in the amygdala have an impaired capacity to evaluate immediate gratifications. Taking both pieces of evidence together, Bechara (2005) constructs a neural theory of willpower. The author distinguishes between an impulsive system (mainly, ventral striatum and amygdala) which processes information about immediate rewards and a reflective system (mainly, ventromedial and dorsolateral prefrontal cortex and anterior cingulate) which processes information about future rewards. These two broadly defined sets of brain structures roughly correspond to our agent and principal (see Bechara (2005, Fig. 1)). Samuel McClure et al. (2004) take the analysis one step further. Based on their fMRI experiments, they argue that the interaction between short-sighted and far-sighted systems provides neuroscientific support for hyperbolic discounting. This view has been recently challenged by Paul Glimcher, Joseph Kable and Kenway Louie (2007).

3. Incentive salience. The importance of impulses and urges in the behavior of emotional an addicted subjects has long been recognized but rarely modelled in economics (Carrillo 2005). The innovative work in neuroscience by Terry Robinson and Kent Berridge (2003) and Berridge (2003) shows that one system mediates the feeling of pleasure and pain (the "liking" system) and a different system mediates the motivation or

incentive to seek pleasure and avoid pain (the "wanting" system). Using pharmacological manipulations, the authors demonstrate that intervention in the mesolimbic dopamine system (MDS) can enhance the willingness of rats to work for food without affecting the benefit of eating it. In a related experiment, subliminal stimuli can alter manifested choices of consumers (wanting decision) without affecting the expected pleasure derived from the commodities (liking outcome). Although, their work is particularly relevant for addiction (see Robinson and Berridge (2003) and the related economic model proposed by Bernheim and Rangel (2004)), this incentive salience mechanism also applies to other impulse-driven choices (Berridge, 2003). The authors acknowledge that wanting and liking interact through an intricate web of brain circuits. They also emphasize the role of the nucleus accumbens and the amygdala in the mediation of wanting, and the role of the prefrontal cortex in overriding MDS-generated impulses (Berridge and Robinson (2003, Fig. 2)). Furthermore, it is suggested that motivational salience can be manifested without conscious awareness.

The combination of evidence about asymmetric information, temporal horizon and incentive salience provides interesting insights. First, the evaluation of alternatives with immediate effects originates in the areas of the brain that we have labelled as impulsive and short-sighted (ventral striatum and amygdala among others). Second, planning, mediation, anticipation of future events, and other high level cognitive functions are located in the areas of the brain that we have labelled as reflective and far-sighted (prefrontal cortex and anterior cingulate among others). Third, the reflective system exerts regulatory control on the impulsive system. At the same time, the impulsive system manages to influence the choices of the reflective system (Miller and Cohen (2001), Bechara (2005)). It should be acknowledged that this review constitutes only a fraction of the current neuroscientific research on the subject. Furthermore, some of these theories have raised serious controversies, which are not discussed here for space considerations. Nonetheless, we argue that taken together they provide support for a brain architecture based on a partly uniformed, forward-looking principal and a better informed, short-sighted, motivationally biased agent.

A last clarification is in order. On the one hand, we advocate a literal interpretation of our dual-system model: the brain is, and therefore should be modelled as, a multi-system structure. On the other hand, the revelation games, incentive contracts and optimization processes are based on the usual 'as if' economic approach. Despite the abstract flavor of the optimal mechanisms, there is a natural way to implement them, which is discussed

in section 2.5.

2 Temporal and informational conflicts in the brain

2.1 The general setting

We consider an individual who lives a finite number of periods $t \in \{1, 2, \ldots, T\}$. At each period t, the individual undertakes two actions, $x_t \in X_t$ and $y_t \in Y_t$. Each action can be pleasant (purchase of commodities, enrollment in leisure activities) or unpleasant (dieting, working). The instantaneous utility of the individual is:

$$U_t(x_t, y_t; \theta_t)$$

where $\theta_t \in \Theta_t$ is a parameter that captures the relative (positive or negative) appeal of the different actions.

Our first brain conflict, namely the *differences in time-horizon*, is modelled in the tradition of Thaler and Shefrin (1981). First, there is one entity, the principal (she) who is cognitive and forward-looking. Second, at each date t there is another entity, agent-t (he) who is impulsive and myopic. Agent-t maximizes his instantaneous utility $U_t(x_t, y_t; \theta_t)$ without any concern for the past or the future. The principal maximizes the sum of utilities of agents in the remaining periods. This temporal conflict of the self has been suggested in several disciplines. Thaler and Shefrin (1981) provide a first formalization in economics under a "Planner and Doer" label. Bechara (2005) refers to the "Reflective and Impulsive" systems in his neurocognitive theory of willpower. In this paper, we adopt a more neutral "Principal and Agent" terminology borrowed from contract theory. Formally, S_t, the intertemporal utility of the principal from the perspective of date t is:

$$S_t = \sum_{s=t}^{T} U_s(x_s, y_s; \theta_s)$$

There are two reasons why we do not impose any exogenous time-preference rate from the principal's viewpoint. First, it sharpens the contrast between principal and agent. Second and most importantly, the choice resulting from the conflicts between brain systems may exhibit a time-preference. Our assumption allows us to identify it as the consequence of such conflicts without any exogenous interference (see section 2.4). In what follows, we assume that the principal controls *at no cost* the actions taken at date t. She may, however, choose to keep an information channel open and be receptive to the signals sent

by agent-t.[6] This formalization captures two basic premises of the relationship between impulse and cognition: the reflective system is ultimately responsible for choices, but the impulsive system can affect these choices (Bechara, 2005). A more detailed discussion about implementation is provided in section 2.5.

Our second brain conflict, the *restriction in the flow of information*, is modelled in the tradition of the contract theory literature. We assume that, even though the principal can impose her preferred actions (x_t, y_t) at each date t, only agent-t knows θ_t, their relative desirability. Such an assumption captures the physiological restrictions brain systems encounter when trying to access information, or the limited conscious awareness of motivations discussed before. This asymmetry of information is problematic for the principal since her optimal decision depends on the parameter θ_t. It is worth emphasizing that our principal and agent are not two localized brain areas. Instead, each system is composed of several brain structures, which play a more or less important role depending on the application. What is key for our analysis is that there are temporal and informational conflicts between these two sets of structures, and that there are no conflicts, information asymmetries, or aggregation problems within each system.

Finally, we introduce scarcity into our model by assuming that actions are linked by an intertemporal constraint:

$$B\left(\{x_s\}_{s=1}^{T}, \{y_s\}_{s=1}^{T}; \{\theta_s\}_{s=1}^{T}\right) \leqslant 0$$

The function $B(\cdot)$ can have different interpretations. It may represent a budget constraint; there is an initial endowment and expenditures in the different goods deplete the budget. Alternatively, if one activity requires income and the other generates it, the constraint may reflect an intertemporal budget balance that must be satisfied between the two. More generally, the function may capture the existence of positive or negative internalities, where current actions affect the utility of future actions (e.g., a meal high in cholesterol and a cigarette provide immediate pleasure but decrease future health, whereas an hour spent at the gym requires effort but improves health).

[6]For the purpose of our model, it can also be assumed that agent-t is in charge of decisions and the principal can costlessly restrict the set of alternatives at his disposal. This is the approach followed for example by Thaler and Shefrin (1981) and Fudenberg and Levine (2006). It is important to note that this alternative formulation where reward circuits are assumed to have control over actions has a weaker neurobiological foundation.

2.2 Consumption and labor under full information

For expositional considerations, the rest of the section focuses on a particular application. Later, we discuss how to modify the analysis in order to capture other situations. At each date t, the individual chooses the amount of pleasant *consumption* $c_t \in C_t = [0, +\infty)$ and unpleasant *labor* $n_t \in N_t = [0, \bar{n}]$. The instantaneous utility is:

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

where $u' > 0$ and $u'' < 0$, and θ_t is the willingness to consume at date t, henceforth referred to as valuation or type. For each unit of labor, the individual obtains one unit of income that can be consumed in any period. Assume a perfect capital market where the individual can save and borrow at the exogenous interest rate r. The intertemporal budget constraint, $B(\cdot)$, takes the following form:

$$\sum_{t=1}^{T} c_t(1+r)^{T-t} \leqslant \sum_{t=1}^{T} n_t(1+r)^{T-t}$$

This formalization differs from the standard life-cycle model with only one decision (consumption) and an exogenous income stream: here, future consumption can be increased 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 with unpleasant activities in a given period.

As a benchmark, consider a two-period horizon with full information. Given that the principal can impose her desired levels of consumption and labor at each period, the preferences, and even the existence of agents, is irrelevant to her. The principal solves program $\mathcal{P}^o$:

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

where $(\mathrm{F_t})$ is the feasibility constraint for c_t and n_t and (BB) is the intertemporal budget constraint. Our first result characterizes the solution when $\bar{n}$ is such that the optimal second-period labor is interior (the proof is trivial and omitted).[7]

[7]Sufficient conditions are $\bar{n} < \big(c_1^o(\underline{\theta})(1+r) + c_2^o(\underline{\theta})\big)/(1+r)$ and $\bar{n} > \big(c_1^o(\overline{\theta})(1+r) + c_2^o(\overline{\theta})\big)/(2+r)$. The analysis can easily be extended to other corner solutions where, for example, $n_1^o < \bar{n}$ and $n_2^o = 0$.

Lemma 1 ***(Full information)*** *The optimal consumption and labor pairs* $(c_t^o(\theta_t), n_t^o(\theta_t))$ *selected by the principal at date* t *when* θ_t *is known are:*

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

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

Since there is a positive net return on savings, it is optimal for the principal to require the highest amount of labor in the first period. Second-period labor is then adjusted to meet the intertemporal constraint. Consumption at date t is proportional to agent-t's valuation. Also, for the same valuation, consumption is higher in period 2 than in period 1 because of the positive return on savings (i.e., $c_2(\theta) > c_1(\theta)$ for all θ). 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. Given that first-period labor is maximal and second-period labor is adjusted to meet the intertemporal budget constraint, consumption levels depend only on valuations. That is, there is no intra-period link between consumption and labor. This result depends on the quasi-linear utility formulation. We adopt this functional form precisely because having no exogenous ties between the variables within each period constitutes an interesting benchmark for comparison.

2.3 Imperfect knowledge of valuation

Suppose now that the principal does not know the true valuation. We assume that valuations are independently drawn from the same continuous distribution over the support $\Theta_t = \Theta = [\underline{\theta}, \overline{\theta}]$ for all t with $0 < \underline{\theta} < \overline{\theta}$, a strictly positive density $f(\cdot)$, and a cumulative distribution function $F(\cdot)$ that satisfies the standard monotone hazard rate conditions: $(F(\theta)/f(\theta))' > 0$ and $((1-F(\theta))/f(\theta))' < 0$. Agent-$t$ learns his current willingness to consume θ_t at the beginning of the period. The principal only knows the distribution from which θ_t is drawn.

Asymmetric information in the brain generates *endogenous constraints on optimal choices*. We wish to underscore the methodological importance of this contribution. As reviewed earlier, there exists a literature where the individual is split into entities that play an intra-period game. However, the starting point of these studies is the existence of an exogenous cost (cost of self-control, cost of exerting willpower, cost of

attention, cost of hot choices) that inevitably leads to trade-offs (fewer resources but better allocation, costly thinking but optimal decision-making, higher current utility but increased likelihood of a future hot mode). The specific way of modelling these costs crucially affects which behaviors can and cannot be rationalized. Unfortunately, given the current knowledge in neuroscience, it is difficult to pinpoint the right assumptions for these functions. We propose a different, more agnostic methodology. Rather than a *cost*, our argument rests on asymmetric information, a *constraint* on decision-making. The principal can then freely design any mechanism she wants in order to promote her favorite actions. This approach, borrowed from the mechanism design literature, is based on more primitive assumptions (conflicts between brain systems) and does not presuppose a specific tradeoff.

With this in mind, we offer a second benchmark for comparison. This benchmark consists of the optimal choices when the principal cannot (or chooses not to) elicit information from the agents. In this case, she precommits to the actions that provide the highest expected utility, that is, she solves the program $\mathcal{P}^{oo}$:

$$\begin{aligned} \mathcal{P}^{oo}: \quad \max_{\{c_1,n_1,c_2,n_2\}} \quad & \int_\Theta \int_\Theta \Big[\theta_1\, u(c_1) - n_1 + \theta_2\, u(c_2) - n_2\Big]\, dF(\theta_1)\, dF(\theta_2) \\ \text{s.t.} \quad & c_t \geqslant 0, \quad n_t \in [0, \bar{n}] \quad \forall\, t \\ & c_1(1+r) + c_2 \leqslant n_1(1+r) + n_2 \end{aligned}$$

Assuming that $\bar{n}$ is such that the optimal second-period labor is interior, the solution is as follows (the proof is again trivial and omitted).

Lemma 2 ***(Asymmetry with no communication)*** *The optimal consumption and labor pairs (c_t, n_t) selected by the principal at date t under asymmetric information and when there is no communication is:*

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

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

The principal cannot set a consumption level that varies with the valuation, and thus ends up choosing an average amount of consumption. Naturally, this is above optimal in low valuation days and below optimal in high valuations days. The individual, nonetheless, works the maximum amount in period 1.

The principal can improve on that solution by deciding to elicit information from the agent. By the very nature of the problem, the principal deals with agent-1 and agent-2 sequentially, so the game is solved by backward induction. At date 2, there is no conflict of preferences between the principal and agent-2 ($S_2 \equiv U_2$). Hence, the choice set of agent-2 does not need to be constrained. Equivalently, agent-2 does not have any incentive to send signals that could mislead the principal about his current valuation. Assuming that agent-1 has consumed and worked (c_1, n_1) and that the weak inequality (BB) has to be satisfied, the levels of consumption and labor in date 2 are identical to those in section 2.2:

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

At date 1, rather than full freedom or full control, the principal relies on an incentive mechanism. More precisely, the principal restricts the choice set of agent-1 to a menu of pairs $\{(c_1(\theta_1), n_1(\theta_1))\}$. Agent-1 can choose any of these pairs or send signals informing the principal which pair he prefers. Applying the revelation principle, this direct mechanism achieves the maximal (second-best) welfare of the principal if it solves the following program $\mathcal{P}^*$:

$$\begin{aligned}
\mathcal{P}^* : \max_{\{(c_1(\theta_1), n_1(\theta_1))\}} \quad & S_1 = \int_\Theta \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) + E_{\theta_2}\Big[\theta_2\, u(c_2^*(\theta_2)) - n_2^*(\theta_2)\Big]\, dF(\theta_1) \\
\text{s.t.} \quad & \theta_1\, u(c_1(\theta_1)) - n_1(\theta_1) \geqslant \theta_1\, u(c_1(\tilde{\theta}_1)) - n_1(\tilde{\theta}_1) \qquad \forall\ \theta_1,\, \tilde{\theta}_1 \qquad \text{(IC)} \\
& c_1(\theta_1) \geqslant 0, \quad n_1(\theta_1) \in [0, \bar{n}] \qquad \text{(F)}
\end{aligned}$$

In $\mathcal{P}^*$, the principal maximizes expected welfare under the feasibility constraint (F), as in program $\mathcal{P}^{oo}$. The solution must also satisfy an incentive compatibility constraint (IC).[8] This latter constraint ensures that agent-1 weakly prefers the pair $(c_1(\theta_1), n_1(\theta_1))$ rather than any other pair $(c_1(\tilde{\theta}_1), n_1(\tilde{\theta}_1))$ with $\tilde{\theta}_1 \neq \theta_1$ when his valuation is θ_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 $\mathcal{P}^*$ characterizes the second-best levels of consumption and labor at date 1 from the principal's viewpoint given the information asymmetry.

Proposition 1 ***(Asymmetric information with temporal conflict)*** *There exists a cutoff θ_1^* $(< \overline{\theta})$ such that the principal restricts the choice set of agent-1 to a menu*

[8] Contrary to standard contract theory problems, this program has no participation constraint. Note, however, that the bounds $c_1 \geqslant 0$ and $n_1 \leqslant \bar{n}$ play a related role in ensuring a minimum utility to the agent. Standard techniques need to be modified to deal with this variation of the problem.

$\{(c_1^*(\theta_1), n_1^*(\theta_1))\}_{\theta_1=\underline{\theta}}^{\theta_1^*}$ *of consumption and labor pairs given by:*[9]

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

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

If $\theta_1 \in [\underline{\theta}, \theta_1^*]$*, agent-1 selects the pair* $(c_1^*(\theta_1), n_1^*(\theta_1))$*. If* $\theta_1 \in (\theta_1^*, \overline{\theta}]$*, agent-1 selects the same pair* $(c_1^*(\theta_1^*), \bar{n})$ *as an agent-1 with valuation* θ_1^**. The principal allows agent-2 any pair of consumption and labor provided that it satisfies* (BB)*. Agent-2 selects:*

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

Proposition 1 shows that neither full delegation nor full control is optimal. For instance, the principal would like agent-1 to consume $c_1^o(\theta_1)$ and work $\bar{n}$ (see Lemma 1) but she cannot tell what the agent's true valuation is. Suppose the principal offers the menu determined in Lemma 1. Because the objective of agent-1 with valuation θ_1 is U_1 rather than S_1, he will pretend to be a type-$\overline{\theta}$ and consume $c_1^o(\overline{\theta})$. In other words, this menu is not incentive compatible. Another option for the principal could be to delegate the choices to agent-1 and let him design his preferred consumption and labor pair. Doing so, however, would be very costly. Since the myopic and selfish agent-1 does not internalize the effect of his choices on agent-2, he would maximize consumption and minimize labor. A third possibility would be to ignore agent-1's information and select the levels of consumption and labor that maximize expected welfare (see Lemma 2). Although an improvement, this would still result in severe inefficiencies. Overall, the best way for the principal to avoid overconsumption is to propose the following rule to agent-1: "Reveal your consumption needs. The higher your reported needs, the higher the consumption you will be allocated but also the higher the amount of work you will provide in exchange." Demanding more work in exchange of more consumption counters agent-1's lack of concern for the future and, at the same time, allows consumption to vary with valuation.

Notice that different valuations do not always translate into different choices, that is, the solution exhibits some pooling. This is the case because agent-1 cannot secure a minimum utility level (see footnote 8). The principal could sort out agent-1's type for

[9] See the appendix for the formal determination of the cutoff θ_1^*.

all θ_1. However, since labor is bounded above by $\bar{n}$, this would require too little work for low valuations and too much consumption for high valuations. She prefers to attenuate these two inefficiencies by granting the same consumption and requiring maximum labor for all valuations above a certain cutoff θ_1^*.

It is important 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) *is not* a result but, instead, a consequence of (BB). By contrast, the self-disciplining rule of working more today to consume more today *is* a result of the asymmetric information model. It is neither first-best nor an ad-hoc restriction. It does not arise when the principal knows the valuations (Lemma 1) or when she disregards the information possessed by agents (Lemma 2). Instead, it emerges as the self-imposed, second-best rule designed by the cognitive system to counter the tendency of the impulsive system to indulge in current satisfaction. Hence, the model provides foundations for behaviors such as: "I will go to this dinner party only if I first exercise for an hour" or "I will eat a slice of this apple pie, but then forego sugar in my coffee."

2.4 The endogenous determination of time preference

In this section, we consider a finite horizon T (> 2). The choices under full information are not qualitatively affected. If $\bar{n}$ is sufficiently large, the consumption granted to agent-s, with $s \in \{1, 2, ..., T\}$, is now:

$$u'(c_s^o(\theta_s)) = \frac{(1+r)^{T-s}}{\theta_s}$$

Labor is maximized in the first τ periods (with $\tau \in \{1, ..., T-1\}$ depending on the value of $\bar{n}$). It is adjusted in period $\tau + 1$ to meet the budget constraint, and there is no labor in periods $\tau + 2$ to T:

$$\sum_{s=1}^{\tau} (1+r)^{T-s}\ \bar{n} + (1+r)^{T-\tau-1}\, n_{\tau+1}^o(\theta_{\tau+1}) = \sum_{s=1}^{T} (1+r)^{T-s}\, c_s^o(\theta_s).$$

Under asymmetric information, the principal does not need to worry about dynamic contracting problems when dealing with each agent, since these have no concern for the future. Also, if types are independently distributed, the valuation revealed by agent-t does not help her improve the contract with agent-$t+1$. Thus, the same principles that apply to the two-period case extend to T periods. Assuming that $\bar{n}$ is such that

the principal can induce sorting in every period (formally, $n_s^*(\underline{\theta}) \geqslant 0$ for all s), we can determine the levels of consumption and labor at each date.

Proposition 2 ***(Extended horizon)*** *At each date $s \in \{1, \ldots, T-1\}$, there exists a cutoff θ_s^* $(< \overline{\theta})$ such that the principal restricts the choice set of agent-s to a menu $\{(c_s^*(\theta_s), n_s^*(\theta_s))\}_{\theta_s=\underline{\theta}}^{\theta_s^*}$ of consumption and labor pairs given by:*[10]

$$u'(c_s^*(\theta_s)) = \frac{(1+r)^{T-s}}{(1+r)^{T-s}\,\theta_s + \left[(1+r)^{T-s} - 1\right]\left(\frac{F(\theta_s)}{f(\theta_s)}\right)}$$

$$n_s^*(\theta_s) = \bar{n} - \left[\overline{\theta}\,u(c_s^*(\overline{\theta})) - \theta_s\,u(c_s^*(\theta_s)) - \int_{\theta_s}^{\overline{\theta}} u(c_s^*(x))dx\right]$$

Agent-s chooses $(c_s^*(\theta_s), n_s^*(\theta_s))$ if $\theta_s \in [\underline{\theta}, \theta_s^*]$ and $(c_s^*(\theta_s^*), \bar{n})$ if $\theta_s \in (\theta_s^*, \overline{\theta}]$. Agent-$T$ is only required to satisfy (BB).

The intraperiod link between consumption and labor is preserved. However, the temporal horizon influences both the levels and the relationship between consumption and labor at each period. This is somewhat expected: even under full information, the number of remaining periods affects the opportunity cost of current consumption and the value of current labor. The novelty is that the amount of extra consumption that the principal needs to grant due to her lack of knowledge of the agent's desires (the informational rents) is also affected by the horizon. Since labor is directly tied to consumption, the amount of extra work also depends on T.

This multi-system approach to intertemporal decision-making allows us to examine a more fundamental question: the *origin of discounting*. In the traditional literature, the role of discounting is to reflect an observed tendency of individuals to prefer the present. The standard model in the absence of discounting is formally equivalent to our model in the case of full information. Therefore, the choice of a patient individual is given by the equation that describes first-best consumption in our model. We can immediately see that, for a given valuation θ, consumption increases over time: $c_{s+1}^o(\theta) > c_s^o(\theta)$. This occurs because the positive interest rate on savings implies a larger opportunity cost of consumption in early periods than in later periods. Since, in practice, we typically observe a preference for the present, it has been necessary to introduce a utility formulation capable of predicting decreasing consumption. The discounted utility formulation,

[10] See the appendix for the formal determination of the cutoff θ_s^*.

introduced by Paul Samuelson (1937) and axiomatized by Tjalling Koopmans (1960), postulates an exogenous rate of impatience and achieves that goal.

The most basic formulation of the discounted utility model assumes, among other things, that discount rates are stationary, intertemporally independent, and constant across activities. Thus, its simplicity and mathematical elegance comes at the expense of realism, as demonstrated in numerous empirical and experimental studies (Shane Frederick, Loewenstein and O'Donoghue 2002). Using insights from psychology, the behavioral economics literature proposes some variations of the model that describe more accurately the dynamic choices of individuals. A prominent example is hyperbolic discounting (Strotz 1956). The main problem is that, whichever variation we adopt, it is always based on some exogenous formulation of time-preferences.

Our model proposes to take one step back. In what follows, we derive the dynamic properties of the consumption path based exclusively on strategic interactions between brain systems –uninformed utilitarian principal versus informed myopic agents– and show that the equilibrium behavior is consistent with observed choices. Thus, our approach allows us to identify the endogenous mechanisms that lead to observed impatience, without relying on any exogenous time-preference parameter.

In order to elicit valuations, the principal has to grant extra consumption. Therefore, the same positive interest rate that makes early consumption to have a higher opportunity cost also implies that early labor is more valuable. This means that, for each unit of labor, the principal is willing to grant more consumption in early periods than in late periods under asymmetric information: $dc_s^*(\theta)/dn_s^*(\theta) > dc_{s+1}^*(\theta)/dn_{s+1}^*(\theta)$. In turn, it implies that, other things being equal, consumption decreases over time: $c_s^*(\theta) > c_{s+1}^*(\theta)$. In other words, for any positive interest rate the informational conflict results in a positive rate of time-preference. Discounting here is derived from the conflicts between brain systems rather than assumed as an intrinsic feature of preferences.

This conclusion can be further developed. Consider an individual with no brain conflict. Assume that period t ($\geqslant 2$) is, from the perspective of period 1, discounted at an exogenous rate $\delta(t-1)$ which, for simplicity, is assumed to satisfy time separability (exponential discounting corresponds to $\delta(t-1) = \delta^{t-1}$). In the absence of commitment to future actions, a simple extension of the first-best consumption level $c_s^o(\theta)$ implies that the optimal consumption at date s under this formulation of discounting, $c_s^\delta(\theta)$, is:

$$u'(c_s^\delta(\theta)) = \delta(T-s)\frac{(1+r)^{T-s}}{\theta}.$$

By equating this consumption to the consumption of an asymmetrically informed principal who puts equal weight on all periods (as described in Proposition 2), we can identify a preference for the present, or degree of impatience, that depends on the intrapersonal information asymmetry. The formulation together with its main properties are summarized in the next proposition.

Proposition 3 ***(Endogenous time preference)*** *Under asymmetric information and given myopic agents and a utilitarian principal, the implicit discount rate is:*

$$\delta(t) = \frac{\theta}{\left[(1+r)^t\right]\theta + \left[(1+r)^t - 1\right]\left(\frac{F(\theta)}{f(\theta)}\right)}$$

Some properties of this function are:

(i) Positive time preference rate: $\delta(t+1) < \delta(t)$ (< 1) *for all* t.

(ii) Decreasing impatience: $\delta(t)/\delta(t-1) < \delta(t+1)/\delta(t)$.

(iii) Steeper discounting the higher the informational rents: as $F(\theta)/f(\theta)$ *increases, both* $\delta(t)$ *and* $\delta(t)/\delta(t-1)$ *decrease.*

The first property, a higher value being attached to close events relative to distant ones, is the most basic finding of studies on discounting.[11] As already discussed, this is the result of larger informational rents (that take the form of increased consumption) being granted in earlier periods in exchange for labor. The second and third properties relate to modern behavioral theories of time-evaluation. Indeed, a period-to-period discount rate that falls monotonically is the defining property of hyperbolic discounting. Although still controversial, this characteristic of time preferences has received substantial support from experimental and empirical research first in psychology and now in economics (Frederick, Loewenstein and O'Donoghue 2002). According to Proposition 3, our brain conflicts may be at the source of this behavioral anomaly. As for the third property, it has also been argued that individuals may not necessarily have a unique discount function (Frederick, Loewenstein and O'Donoghue 2002). Preliminary evidence in Loewenstein et al. (2001) suggests that people exhibit different rates of time-preference for different categories of activities (e.g., repetitive tasks versus viscerally driven behaviors). One can argue that idiosyncratic preference shocks are less predictable, and therefore informational rents are more important, in settings subject to impulsive reactions (indulging a vice) than

[11] There are, however, examples of negative time preferences as illustrated, for example, in Loewenstein and Prelec (1991).

in recurrent tasks (flossing one's teeth). Under this assumption, our model predicts a steeper discounting in the former than in the latter category of activities.

2.5 Implementation

The previous analysis raises the question of how to map our abstract mechanism into a neural theory. To answer this, we first need to determine which neural circuitries are implicated in the evaluation of alternatives (willingness to consume, displeasure of labor, value of income). There is solid evidence that the ventral striatum (nucleus accumbens, ventral caudate and ventral putamen) is part of the circuitry involved in the processing of primary rewards such as food or drugs (Berns et al., 2001). Recent fMRI studies show that it is also involved in the evaluation of aversive events such as noxious thermal shocks (Lino Becerra et al., 2001) and cutaneous electrical stimulations (James Jensen et al., 2003). Perhaps more surprisingly, the striatum is also implicated in incentive-driven rewards like monetary gains and losses (Brian Knutson et al. (2000), Mauricio Delgado et al. (2000)). Taken together, this body of research suggests that similar neural networks are responsible for encoding different types of values: goods with hedonic properties, negative stimuli and even pure conditioned rewards. As summarized by Rebecca Elliott et al. (2003, p. 303): "it is clear that the neuronal substrates of financial reinforcement overlap extensively with regions responding to primary reinforcers, such as food." In terms of our model, the same agent is likely to be in charge of evaluating enjoyable and displeasureable activities.

Once this is established, we can ask ourselves how the disciplining rule described in the previous sections can be implemented in practice. Unfortunately, to the best of our knowledge no work has been designed to address this question. We can combine evidence from different studies to suggest a possible mechanism. However, the argument is necessarily speculative. First, the brain structures in the cortical systems (our principal), who are ultimately responsible for choices and have a mental representation of the future consequences of current actions, 'commit' to a subset of choice pairs. This can be achieved, for example, by limiting the amount of signals coming from lower systems that are processed.[12] Second, the systems that encode value (our agent) receive anticipatory information about the value of each good or activity. The key, as discussed

[12] An information censoring of this type is discussed in Bechara's (2005) neurocognitive theory of willpower: "Another mechanism of impulse control is the ability to resist the intrusion of information that is unwanted" (p. 1460).

above, is that overlapping systems are activated for rewards of vastly different nature. Moreover, according to Read Montague and Berns (2002), the information about these disparate rewards (money, food, sex, work) is accumulated and converted into a common scale or 'neural currency', which is then used to compare alternatives. This aggregation process occurs in the orbitofrontal-striatal circuit. Third, once the relative importance of rewards is evaluated, the striatal system 'communicates' its preferred pair to the motor cortex (Knutson et al., 2000). This may be done by sending some neuronal signals carrying information about the desirability of x and some other signals carrying information about the desirability of y. If all signals were processed, the striatal would overstate the positive (negative) value of any pleasant (unpleasant) activity. Because the amount of information processed is restricted (see the first point), it is in the agent's best interest to carefully select the *relative number* of signals in favor of each alternative.

3 Some implication for choice over time

3.1 Choice bracketing and expense tracking

Studies in marketing and psychology show that consumers often set budgets for narrowly defined categories (clothing, entertainment, food) and track expenses against budgets (Thaler (1985), Itamar Simonson (1990)). The cost of narrow choice bracketing is obvious: it forces consumers to perform local rather than global maximizations. The benefit is less clear. Read, Loewenstein and Rabin (1999) suggest that narrow bracketing requires less involving calculations and can be used as an effective self-disciplining mechanism to avoid excesses. However, we are not aware of any model that formalizes this or any other potential advantage. The argument seems intuitive, but not fully satisfactory. First, nothing prevents a broad bracketing consumer from mimicking a narrow bracketing one. Second and more importantly, the experiments of Chip Heath and Jack Soll (1996) demonstrate that a narrow definition of categories leads people to underconsume some goods and *overconsume some others*.

We propose a different rationale for this behavior.[13] Following the general model described in section 2.1, consider an individual who intertemporally allocates a fixed initial income k between two classes of goods, clothing ($x_t \geqslant 0$) and entertainment ($y_t \geqslant 0$). The principal can select her desired composition of expenditures but ignores

[13]For the sake of brevity, we only describe a sketch of the model. Detailed proofs of the arguments for the different cases are available upon request.

the relative willingness θ_t of agent-t to consume each good. Formally, the instantaneous utility is:

$$U_t(x_t, y_t; \theta_t) = \theta_t u(x_t) + y_t.$$

The intertemporal budget constraint, $B(\cdot)$, is:

$$\sum_{t=1}^{T} (x_t + p\, y_t)(1+r)^{T-t} \leqslant k(1+r)^{T-1},$$

where 1 and p are the unitary prices of goods x and y. If decisions are delegated, agent-1 chooses the optimal allocation across goods in period 1, but he exhausts the budget. Following Lemma 2, the principal can also limit the per-period budget of the tempting good to its expected optimal level (as in the precommitment rules developed by Thaler and Shefrin (1981)). However, as demonstrated in Proposition 1, the principal can do better by imposing a per-period negative relationship between expenditures in each category. The strategy does not lead to first-best optimality. Nevertheless, it requires a simple rule of behavior and enables the person to achieve some self-discipline, the advantages of narrow bracketing described in the literature.[14] Furthermore, if valuations for the goods are independent, this self-imposed negative correlation of expenditures will generate, on average, overconsumption of one good and underconsumption of the other. Thus, it reconciles the self-control motive for mental accounting emphasized by Thaler (1985) with the simultaneous feeling of wealth and poverty described in Heath and Soll (1996).

A similar argument can rationalize the tendency of self-employed individuals (fishermen, salesmen, writers) to work longer hours on less productive days. Consider the case of New York City cabdrivers. Assume that the principal does not have access to the information regarding the difficulty to earn money, and that the agent dislikes working. Delegation results in shirking. The principal can achieve some self-discipline and a second-best allocation of time by arbitrarily dividing the day into several subperiods (e.g., morning and afternoon). Formally, denote by l_t^m and l_t^a labor in the morning and labor in the afternoon, and assume that one unit of labor translates into one unit of earnings. Each day, the agent maximizes $U_t(l_t^m, l_t^a; \theta_t) = -\theta_t \psi^m(l_t^m) - \psi^a(l_t^a)$ where $\psi^m(\cdot)$ and $\psi^a(\cdot)$ represent the disutilities of labor, and θ_t captures a shock in the relative difficulty to earn

[14] If the optimal (concave) relationship between expenditures in the two commodities is cognitively too difficult to implement, the individual may resort to a simpler (linear) relationship at a small extra utility loss (we thank an anonymous referee for raising this issue).

money. The principal cares about the utility of all agents. The budget constraint, $B(\cdot)$, is:

$$\sum_{t=1}^{T}(l_t^m + l_t^a)(1+r)^{T-t} \geqslant C\sum_{t=1}^{T}(1+r)^{T-t},$$

where C represents the daily consumption of the agent. In this case, the principal proposes an incentive mechanism where labor in the afternoon is inversely related to earnings in the morning: the agent is allowed to work less in the afternoon if earnings in the morning are higher. An intrapersonal contract of this type can partly explain the puzzling negative elasticity of wages and hours of work documented by Camerer et al. (1997).

3.2 Life-Cycle theory

The life-cycle model provides a framework to study intertemporal consumption. This theory makes several predictions. First, holding intertemporal levels constant, the dynamics of income accumulation should not affect the dynamics of consumption. Second, the propensity to consume current income should be independent of its source. Third, if discretionary savings are positive, then an increase in pension savings should not affect total savings. Empirical analyses (e.g., Robert Hall and Frederick Mishkin (1982)) suggest that people behave quite differently: the propensity to consume strongly depends on current income, on the source of wealth and on the level of mandatory savings (see Shefrin and Thaler (1988) and Thaler (1990) for reviews of the empirical anomalies). Several theories have been proposed to explain these differences. They include bequest motives, capital market imperfections, changing preferences, self-control problems, and mental accounting rules. Our approach may help explain some of the links between income and consumption in a unified framework.

First, our model predicts that, controlling for total wealth, consumption tracks earned income. The intuition is simple. Assume that either the pleasure of consumption or the disutility of labor varies from period to period and is only known to the agent. The principal achieves self-discipline with the rule work more to consume more. Consumption is above its first-best level, but excesses are mitigated. By contrast, if the individual is retired or unemployed, this compensatory mechanism cannot be used. To avoid maximum consumption, the principal must impose no fluctuations, that is, the consumption chosen by an average type under full information (see Lemma 2). Note that our theory predicts not only lower average levels but also smaller fluctuations in consumption during retirement or unemployment. We are not aware of any existing test of this hypothesis.

Second, the source of wealth affects the propensity to consume. Consumption is granted in exchange of *costly* effort. Therefore, as income is obtained from a less costly source (capital gain, windfall, income borrowed against future labor), the principal loses the ability of using this tool to elicit valuations. The evidence provides mixed support for this prediction. On the one hand, income which is more costly to obtain is spent in larger proportions: the propensity to consume regular income is greater than the propensity to consume a bonus which is itself greater than the propensity to consume a capital gain (Shefrin and Thaler, 1988). This finding is consistent with our theory. On the other hand, consumption is excessively correlated with most income changes, including windfalls. Our theory cannot explain this finer result.

A third and more subtle prediction relates to the effect of mandatory savings on total savings. Note that $dn_1^*(\theta_1)/dc_1^*(\theta_1) > 0$ and $d^2n_1^*(\theta_1)/dc_1^*(\theta_1)^2 < 0$. This means that a higher valuation agent consumes a bigger fraction of his earned income. Therefore, a mandatory savings rate (e.g., a pension plan) constrains only the consumption choices of agents whose valuation is above a certain cutoff $\tilde{\theta}$. Interestingly, a mandatory savings rate relaxes the incentive problem for high valuation agents and, given the positive rate of return, it is optimal to increase their labor in exchange of this reduced consumption. In turn, it is also optimal to shift upwards the labor of agents with valuations below $\tilde{\theta}$, which results in increasing their savings also. This imperfect substitutability between mandatory and discretionary savings captures another behavioral anomaly documented in the literature.

4 Incentive and informational conflicts in the brain

4.1 The general setting

Temptation puts the individual in a state of mind where activities that provide a moderate objective satisfaction suddenly become irresistible. Salient motivations or impulsive urges may be pathological (eating disorder, bipolar disorder, or obsessive-compulsive disorder). They are most prevalent for addictive substances (Robinson and Berridge (2003) and Bernheim and Rangel (2004)). However, the recent evidence from neuroscience suggests that different systems mediate the feeling of pleasure (liking) and the motivation to seek pleasure (wanting). Furthermore, discrepancies may manifest also for regular goods (Berridge, 2003). In this section, we incorporate our third conflict, namely the dichotomy between liking versus wanting, in our dual-system model of the brain. To better focus on

incentive salience and informational asymmetry, we abstract from the temporal conflict. More precisely, the individual engages in two activities, x and y, during one period. The true instantaneous payoff of the individual is:

$$U(x, y; \theta) = \theta\, u(x) + v(y)$$

where θ represents the valuation of the more tempting good x relative to the less tempting (or non-tempting) good y. We assume that $\theta \in \Theta = [\underline{\theta}, \overline{\theta}]$ and that its c.d.f. $F(\theta)$ satisfies the same hazard rate conditions as in section 2. Function $U(\cdot)$ is the utility representation of the "liking" system (the principal), which captures how consumption of the different goods does affect welfare. However, what motivates the individual to consume is:

$$W(x, y; \theta) = \theta\, w(x) + v(y)$$

Function $W(\cdot)$ is the utility representation of the "wanting" system (the agent), which captures how perceived welfare and choices are biased by visceral influences. We assume that $u(0) = 0$, $u'(x) > 0$, $u''(x) < 0$ and $w(0) = 0$, $w'(x) > 0$, $w''(x) < 0$: both principal and agent find good x enjoyable, although they might disagree on its contribution to welfare. In this one-period problem, the scarcity or budget constraint, $B(\cdot)$, takes the following expression:

$$x - r(y) \leqslant 0$$

The utility of the principal and the budget constraint of the consumption and labor model studied in section 2.2 correspond to $v(y) = -y$ and $r(y) = y$, with the variables c and n being replaced by x and y respectively. The choice bracketing application with two pleasurable goods briefly presented in section 3.1 corresponds to $v(y) = y$ and $r(y) = k - py$. We will assume that either $v'(y) > 0$ and $r'(y) < 0$ for all y (activity y is pleasant but tightens the budget constraint) or $v'(y) < 0$ and $r'(y) > 0$ for all y (activity y is unpleasant but softens the budget constraint). Let us call $\mathcal{U}$ and $\mathcal{W}$ the optimization programs of the principal and the agent when θ is common knowledge:

$$\mathcal{U}: \max_{x,y} \ \theta\, u(x) + v(y) \quad \text{s.t.} \quad x \leqslant r(y) \qquad \text{and} \qquad \mathcal{W}: \max_{x,y} \ \theta\, w(x) + v(y) \quad \text{s.t.} \quad x \leqslant r(y)$$

To ensure concavity of these optimization programs, we make the following assumption.

Assumption 1 *The utility of the principal and the agent satisfy:*[15]

[15]Note that, if $r(y)$ is linear, a sufficient condition for assumption 1 to hold is $v''(y) \leqslant 0$.

$$\theta u''(z) + v''(r^{-1}(z))[r^{-1'}(z)]^2 + v'(r^{-1}(z))r^{-1''}(z) \leqslant 0 \quad \forall\, z,\, \theta$$
$$\theta w''(z) + v''(r^{-1}(z))[r^{-1'}(z)]^2 + v'(r^{-1}(z))r^{-1''}(z) \leqslant 0 \quad \forall\, z,\, \theta$$

Denote by $(x^F(\theta), y^F(\theta))$ and $(x^D(\theta), y^D(\theta))$ the optimal choices of principal (first-best) and agent (full delegation). These pairs maximize $\mathcal{U}$ and $\mathcal{W}$, respectively:

$$\theta u'(x^F(\theta)) + v'(r^{-1}(x^F(\theta)))r^{-1'}(x^F(\theta)) = 0 \quad \text{and} \quad y^F(\theta) = r^{-1}(x^F(\theta))$$

$$\theta w'(x^D(\theta)) + v'(r^{-1}(x^D(\theta)))r^{-1'}(x^D(\theta)) = 0 \quad \text{and} \quad y^D(\theta) = r^{-1}(x^D(\theta))$$

In both cases, the budget constraint binds since valuable resources should not be wasted. Differentiating the first-order conditions, we get $dx^F/d\theta > 0$ and $dx^D/d\theta > 0$: a higher valuation translates into a greater consumption under first-best and delegation. Incentive salience states that the agent wants to consume an amount of the tempting good x which is considered excessive by the principal, that is, $x^D(\theta) > x^F(\theta)$ for all θ. The following assumption ensures that this inequality holds.[16]

Assumption 2 $u'(x) < w'(x) \;\; \forall\, x.$

Given $u(0) = 0$ and $w(0) = 0$, assumption 2 also implies that $u(x) < w(x)$ for all x. Last, we denote by $T(\cdot)$ the function that transforms the utility of the agent for good x into the utility of the principal:

$$u(x) = T(w(x))$$

where $T(z) > 0$ and $T'(z) > 0$ for all z. Given assumption 2, $T'(z) < 1$ for all z.

4.2 Incentive salience and optimal delegation of choices

As in section 2, the principal maximizes welfare. Unlike before, the conflict is due to the agent being subject to urges that affect perceived utility ($W(\cdot) \neq U(\cdot)$). Under complete information, biased motivations are irrelevant since the principal can impose her optimal pair of choices $(x^F(\theta), y^F(\theta))$. Under incomplete information, full delegation results in excessive consumption of the tempting good. To combat this tendency, the principal must design a revelation mechanism. Interestingly, the options offered under

[16] $u'(x) < w'(x) \;\Rightarrow\; 0 = \theta\, u'(x^F) + v'(r^{-1}(x^F))r^{-1'}(x^F) < \theta\, w'(x^F) + v'(r^{-1}(x^F))r^{-1'}(x^F) \;\Rightarrow\; x^F < x^D$.

incentive salience are quite different than under temporal conflict. The principal solves the following program $\mathcal{U}_{\text{AI}}$:

$$\mathcal{U}_{\text{AI}}: \max_{\{(x(\theta),y(\theta))\}} \int_{\underline{\theta}}^{\overline{\theta}} \Big[\theta\, u(x(\theta)) + v(y(\theta))\Big] dF(\theta)$$

$$\text{s.t.} \quad \theta\, w(x(\theta)) + v(y(\theta)) \geqslant \theta\, w(x(\tilde{\theta})) + v(y(\tilde{\theta})) \qquad \forall\ \theta,\, \tilde{\theta} \qquad (\hat{\text{IC}})$$

$$x(\theta) \leqslant r(y(\theta)) \qquad (\hat{\text{BB}})$$

The solution $(\hat{x}(\theta), \hat{y}(\theta))$ to program $\mathcal{U}_{\text{AI}}$ characterizes the constrained optimum that the cognitive system can achieve given the private information and biased motivation of the affective system.

Proposition 4 ***(Asymmetric information with incentive salience)***

When $T''(z) \leqslant 0$, the principal sets a consumption cap $\bar{x}$ and requires ($\hat{\text{BB}}$). *Given this rule, there exists a valuation $\hat{\theta}$ such that the agent chooses his optimal pair $(x^D(\theta), y^D(\theta))$ if $\theta < \hat{\theta}$ and the optimal pair $(x^D(\hat{\theta}), y^D(\hat{\theta}))$ of an agent with valuation $\hat{\theta}$ if $\theta \geqslant \hat{\theta}$.*[17]

When $T''(z) > 0$, there exist n $(\geqslant 2)$ subintervals such that:
$\hat{x}(\theta) = x^D(\theta)$ *for all* $\theta \in [\underline{\theta}, \theta_1] \cup [\theta_2, \theta_3] \cup \ldots \cup [\theta_{n-2}, \theta_{n-1}]$;
$\hat{x}(\theta) = x^D(\theta_1)\, \forall \theta \in (\theta_1, \theta_2)$, $\hat{x}(\theta) = x^D(\theta_3)\, \forall \theta \in (\theta_3, \theta_4)$,..., $\hat{x}(\theta) = x^D(\theta_{n-1})\, \forall \theta \in (\theta_{n-1}, \overline{\theta}]$.
If $n > 2$, then resources are wasted (i.e., $x(\theta) < r(y(\theta))$) for all valuations $\theta > \theta_2$.

Contrary to Proposition 1 where intervention was sophisticated and intrusive, the principal now follows a simple rule-of-thumb. Condition $T'' \leqslant 0$ together with assumption 2 implies that $u''(x) < w''(x)$: the marginal disagreement between the principal and the agent increases with the level of consumption, and therefore with the valuation of the tempting good. The cost of letting the agent get away with his desired consumption of x is small as long as his valuation is low. When the valuation exceeds a certain threshold $\hat{\theta}$, 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 against temptation is the principal saying "as long as you don't abuse, you can do whatever you want." Given this rule, the agent makes sure the budget constraint is always binding ($\hat{x}(\theta) = r(\hat{y}(\theta))$), so resources are never wasted.[18]

For the reader familiar with incentive theory, this form of contract should be intriguing. For the sake of exposition, suppose that y is unpleasant ($v'(y) < 0$) and $r(y)$ is

[17] See the appendix for the formal determination of $\bar{x}$. There is also a limit case discussed in the proof where $\bar{x} \leqslant x^D(\underline{\theta})$ and therefore $\hat{x}(\theta)$ is constant for all $\theta \in [\underline{\theta}, \overline{\theta}]$.

[18] Instead of a consumption cap on x, the principal can equivalently set a consumption floor on y.

linear (so the constraint is $x \leqslant y$). The intuition behind the technical aspect of this result is a consequence of the three tools that the principal can use to satisfy incentive compatibility. First and trivially, the principal can let the agent choose the pair he wants. Second, she can force all types of agents to make the same pooling choice. Third, she can optimally select the (monotone) relation between x and y that induces self-selection. In standard problems, incentive compatibility is ensured via the third criterion or a combination of the second and third criteria. By contrast, in our setting, there is a *tension between inducing self-selection and managing resources*. On the one hand, self-selection requires the indifference curves that relate x and y to be increasing and convex. On the other hand, a binding budget constraint requires a linear relation between x and y. This immediately implies that if the principal wants to induce self-selection, she must waste resources. The idea is illustrated in Figure 1a: to preserve convexity of the relation between x and y, the agent consumes $x^F(\theta)$ and all types except $\tilde{\theta}$ engage in excessive y (the slanted area represents the amount of wasted resources, $x(\theta) < y(\theta)$). The other alternative for the principal is to leave full freedom to the agent, in which case $\hat{x}(\theta) = x^D(\theta)$ and $\hat{y}(\theta) = x^D(\theta)$. By definition and as illustrated in Figure 1b (full line), this also results in overconsumption of the tempting good relative to the first-best option (dotted line) but, at least, resources are not wasted. Because overconsumption is especially severe for high-valuation types, the principal finds it optimal to delegate choices and limit the inefficiency of overconsumption by constraining all agents above a certain valuation $\hat{\theta}$ (dashed line).

[Figures 1a and 1b here]

This simple rule has other implications. Keeping the consumption and labor interpretation, it follows that the individual will incur *excesses in both the pleasant and the unpleasant activities*: the principal indulges extra consumption ($x^D(\theta) > x^F(\theta)$) but requires extra work ($y^D(\theta) > y^F(\theta)$). While self-control problems can explain overconsumption and strict rule setting can explain overwork, it is usually difficult to find reasons that explain both types of excesses at the same time. One can also think of the conflict in terms of morality. The principal has a constrained willingness to engage in pleasurable activities that are socially harmful or unaccepted. The agent does not share this high-order moral disposition. Rather than imposing self-discipline for all valuations, our result shows that the principal finds it optimal to simply limit the maximum amount

of the pleasurable activity that the agent is allowed to enjoy.[19] Finally, we can apply this mechanism to a 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 tempting activity, and the parent internalizes only partly his preferences. In such a situation, full delegation of choices up to a point and firm intervention thereafter is the parent's second-best optimal strategy.

What happens when $T'' > 0$? The conflict between the principal and the agent can be either increasing or decreasing in consumption. When the conflict is increasing, we obtain the same insights as before: $n = 2$, so there is delegation for all $\theta \in [\underline{\theta}, \theta_1]$ and pooling (or identical consumption) for all $\theta \in (\theta_1, \overline{\theta}]$. When the conflict is decreasing or non-monotonic, the regions of delegation and pooling as a function of θ alternate. The optimal consumption path of the tempting good is illustrated in Figure 2.

[Figure 2 here]

By allowing identical consumption of x to all types in an interval (say, (θ_{i-1}, θ_i)), the principal moderates excesses. However, delegation in the next interval $[\theta_i, \theta_{i+1}]$ becomes problematic: an agent below but close to θ_i will want to pick the contract of a type-θ_i. To avoid mimicking, the principal must ensure that utility is continuous in valuation. This is achieved by imposing a lump sum change in the other good y to all agents with type θ_i and above. Since the extra change in y exceeds the strict needs to satisfy the budget balance, the constraint ($\hat{\text{BB}}$) becomes slack. Overall, the decision to intervene is governed by the following tradeoff: a longer pooling interval limits overconsumption of the tempting good but requires a bigger jump in consumption at the boundary, and therefore a larger waste of resources to ensure incentive-compatibility. Finally, note that all contractual regimes in $\mathcal{U}_{\text{AI}}$ are characterized by either delegation or pooling, but never by self-selection as in typical mechanism design problems.

We wish to emphasize that the principal implements different incentive mechanisms under temporal and temptation conflicts because the tradeoffs are different. Under temporal conflict, excessive consumption of x has a high cost as it implies that fewer resources are left for the future. By contrast, meeting the budget constraint is not essential since the accumulated resources can be used in the following period(s). Under incentive conflict, the allocation of resources between periods is not an issue, but meeting the budget

[19] See Bénabou and Tirole (2004) for an explanation of compulsive behavior based on hyperbolic discounting, and Rabin (1995) for a different view on the effect of moral preferences and moral constraints on behavior.

constraint is important because unused resources are forever lost.

Finally, program $\mathcal{U}_{\text{AI}}$ is technically very similar to Amador, Werning and Angeletos (2006), where activities are consumption at dates 1 and 2 and the disagreement results from hyperbolic discounting, rather than incentive salience. Their setting coincides with our model under linear conflict ($T'' = 0$). Both papers prove the optimality of a consumption cap rule (or, in their case, a savings threshold rule) under monotone hazard rate and linear conflict.[20] Their paper relaxes the monotone hazard rate assumption whereas our paper relaxes conflict linearity. Under either generalization (but for different reasons), wasting resources may become part of the principal's optimal strategy.

4.3 An example: linear conflict

Consider the special case of a linear conflict between the wanting and liking systems. Formally, let $w(x) = \alpha\, u(x)$ with $\alpha > 1$, so $T''(z) = 0$. Applying Proposition 4 to this particular conflict, we obtain the following result.

Proposition 5 *Under a linear incentive conflict, choices are $(x^D(\theta), y^D(\theta))$ for all $\theta \leqslant \theta_l$ and $(x^D(\theta_l), y^D(\theta_l))$ for all $\theta > \theta_l$, where θ_l is such that $\alpha\, \theta_l = E[\theta \mid \theta > \theta_l]$.*

For a given valuation, the agent is less likely to make free decisions when the conflict is high and when the willingness to consume is drawn from a less favorable distribution.

Fix the utility of the principal $u(x)$. As the impulsive urges become more pronounced (α larger), the gap between the optimal choice of the principal and the motivations of the agent increases, so the former needs to control the latter more tightly. This results in a higher probability of intervention ($\partial\theta_l/\partial\alpha < 0$), as illustrated in Figure 3.

[Figure 3 here]

This also means that more intransigent rules reflect a stronger conflict. Note that $\theta_l(\alpha) < \overline{\theta}$ for all $\alpha > 1$ and $\lim_{\alpha\to 1} \theta_l(\alpha) = \overline{\theta}$: the principal intervenes as soon as there is a difference between true and perceived utility, even if it is minimal. Also, $\theta_l = \underline{\theta}$ for all $\alpha > E[\theta]/\underline{\theta}$: if the bias is sufficiently important the principal imposes the same action for all valuations. Finally, one may argue that the wanting system learns the preferences of the liking system over time or that visceral impulses are better controlled with age and experience (see e.g. the construction of preferences argument discussed previously).

[20] See below for an analytical characterization of this special case and some comparative statics.

Either way, if α moves closer to 1 over time, the incentive scheme shifts towards a more lenient intervention.

The distribution of valuations also affects intervention. Suppose that θ can be drawn from $F(\theta)$ or $G(\theta)$, where $G(\theta)$ stochastically dominates $F(\theta)$, that is, $F(\theta) > G(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta})$. We know that the optimal scheme balances the costs of overconsumption with the costs of pooling. For a given threshold θ_l, consumption is more likely to be restrictive if the distribution is more favorable. In order to avoid an excessive intervention, the principal then becomes more lenient when valuations are more likely to be high.

5 Concluding remarks

The theory of organizations has a long tradition in modelling the firm as a nexus of agents with incentive problems, informational asymmetries, restricted communication, etc. Based on recent neuroscience research, this paper argues that individual decision-making should be studied from that same multi-system perspective and proposes a step in that direction (Brocas and Carrillo (forthcoming) discuss in more detail some advantages of this "neuroeconomic theory" methodology). Other studies have implicitly followed a similar approach. A main difference is that the literature has always focused on automatic processes versus rational optimization whereas we exploit different neuromechanisms: the cognitive inaccessibility to our motivations and the presence of salient motivations.

Some readers may resist the idea of brain modularity. Yet, conflicts between brain systems have been amply demonstrated and are now mainstream in some areas of neuroscience research, such as memory (Poldrack and Rodriguez, 2004) or information processing (Miller and Cohen, 2001). Recent studies even suggest that some systems act as conflict mediators (William Gehring et al. (1993), John Kerns et al. (2004)). Biologists, neuroscientists and psychologists have proposed different evolutionary theories to explain a brain architecture composed of multiple, interacting systems. For example, Richard Dawkins (1976) argues that selection operates at the gene not at the individual level. John Tooby and Leda Cosmides (1992) claim that, in a changing environment, internal conflicts are often a remnant of past evolution. More recently, Adi Livnat and Nicholas Pippenger (2006) show that under some reasonable physiological limitations, the development of modules with conflicting objectives may result in improved outcomes. This last argument should not be too surprising. We know that in competitive environments and given some organizational constraints (bounded resources, restricted channels of commu-

nications), decentralized firms may outperform centralized ones. Since the brain is also subject to all sorts of physiological constraints, it seems reasonable to think that a similar argument could be applied here.

Our model may be extended in several dimensions. We can introduce correlated valuations (or learning over time, as in the construction of preferences approach) and attenuate the conflict by assuming that agents have a positive concern for future returns. This creates a self-signaling problem different from that in Bodner and Prelec (2003) and Bénabou and Tirole (2004): agents require extra rents to reveal their information since that knowledge is subsequently used by the principal to their own detriment (the ratchet effect). We can also allow agents to invest resources that increase their productivity of labor. It may also be interesting to test empirically or experimentally some behavioral implications of our theory. Results of special relevance in our model are: (i) the use of narrow choice bracketing as a self-disciplining device to overcome myopic behavior; (ii) the lower fluctuation in consumption when the individual does not have access to labor; and (iii) the differences in discount rates for categories of activities that are subject to different degrees of idiosyncratic preference shocks.

As a final note, we would like to stress the importance of collaborative ventures between neuroscientists and economists. On the one end, experiments in neuroscience provide invaluable information to economic theorists about how to build better organizational models of the brain. On the other end, theoretical models of decision-making processes can help experimental neuroscientists determine which hypotheses about the architecture of the brain deserve testing priority. Although it is far too early for an assessment, this methodology may eventually result in a new approach to economic decision-making, moving from a single-unit formulation with a centralized decision-maker to a multi-unit formulation with strategic interactions.

Appendix

Appendix A. Proof of Propositions 1 and 2. The principal's objective function at date t is:

$$S_t = E_{\theta_t}\Big[\theta_t\, u(c_t(\theta_t)) - n_t(\theta_t)\Big] + \sum_{\tau=t+1}^{T} E_{\theta_\tau}\Big[\theta_\tau\, u(c^*_\tau(\theta_\tau)) - n^*_\tau(\theta_\tau)\Big]$$

where $c^*_\tau(\theta_\tau)$ and $n^*_\tau(\theta_\tau)$ are anticipated future levels. Agent-t only cares about choices at t. His utility when his valuation is θ_t and he chooses the pair $(c_t(\tilde\theta_t), n_t(\tilde\theta_t))$ is:

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

Incentive Compatibility. The mechanism offered by the principal is incentive compatible if and only if $U_t(\theta_t, \theta_t) \geqslant U_t(\theta_t, \tilde\theta_t) \;\; \forall\, \theta_t,\, \tilde\theta_t$. Let $U_t(\theta_t) \equiv U_t(\theta_t, \theta_t)$. The two necessary and sufficient conditions for incentive compatibility at date t are:[21]

$$\dot U_t(\theta_t) = u(c_t(\theta_t)) \qquad (\mathrm{IC}_1)_\mathrm{t}$$

$$\dot c_t(\theta_t)\frac{\partial U_t}{\partial n_t}\left[\frac{\partial}{\partial \theta_t}\left(\frac{\partial U_t/\partial c_t}{\partial U_t/\partial n_t}\right)\right] \geqslant 0 \quad \Rightarrow \quad \dot c_t(\theta_t) \geqslant 0 \qquad (\mathrm{IC}_2)_\mathrm{t}$$

Feasibility. Labor $n_t(\theta_t)$ must lie in $[0, \bar n]$ and consumption must be positive, that is:

$$U_t(\theta_t) \geqslant \theta_t\, u(c_t(\theta_t)) - \bar n \equiv B^l(\theta_t) \qquad (\mathrm{FL}_1)_\mathrm{t}$$

$$U_t(\theta_t) \leqslant \theta_t\, u(c_t(\theta_t)) \equiv B^u(\theta_t) \qquad (\mathrm{FL}_2)_\mathrm{t}$$

$$c_t(\theta_t) \geqslant 0 \qquad (\mathrm{FC})_\mathrm{t}$$

Budget. At date t, the individual inherits (positive or negative) saving s_{t-1}, consumes c_t, works n_t and leaves (positive or negative) saving s_t for the next period. Since resources can a priori be thrown away, the following budget constraint inequality must hold:

$$s_{t-1}(1+r) + n_t(\theta_t) \geqslant c_t(\theta_t) + s_t \qquad (\mathrm{B})_\mathrm{t}$$

with $s_0 = 0$ (no initial resources) and $s_T \geqslant 0$ (no deficit at the end of the last period).

Program. The objective function of the principal at date t can thus be reduced to the maximization of S_t subject to $(\mathrm{IC}_1)_\mathrm{t}$, $(\mathrm{IC}_2)_\mathrm{t}$, $(\mathrm{FL}_1)_\mathrm{t}$, $(\mathrm{FL}_2)_\mathrm{t}$, $(\mathrm{FC})_\mathrm{t}$, $(\mathrm{B})_\mathrm{t}$.

Period T. There is no conflict between principal and agent-T, so $(\mathrm{IC}_1)_\mathrm{T}$ and $(\mathrm{IC}_2)_\mathrm{T}$ trivially hold. Savings at T are wasted, so $s_T = 0$. Ignoring feasibility, maximization of S_T s.t. $(\mathrm{B})_\mathrm{T}$ implies $c^*_T(\theta_T) = c^o_T(\theta_T)$ and $n^*_T(\theta_T) = c^*_T(\theta_T) - s_{T-1}(1+r)$. We will assume that $\bar n$ is such that $n^*_T(\theta_T) \in [0, \bar n]$ for all θ_T.

[21] Techniques are standard (see e.g. Fudenberg and Tirole (1991, ch. 7)) so the proof is omitted.

No waste of resources. Given, $(c_T^*(\theta_T), n_T^*(\theta_T), s_T)$, we have that at $T-1$:

$$S_{T-1} = E_{\theta_{T-1}}\Big[\theta_{T-1}u(c_{T-1}(\theta_{T-1}))-n_{T-1}(\theta_{T-1})\Big]+E_{\theta_T}\Big[\theta_T u(c_T^*(\theta_T))-c_T^*(\theta_T)\Big]+s_{T-1}(1+r)$$

Since S_{T-1} is increasing in s_{T-1}, then $(\mathrm{B})_{\mathrm{T}-1}$ is binding. Suppose now that $(\mathrm{B})_{\mathrm{t}+1}$ to $(\mathrm{B})_{\mathrm{T}-1}$ are binding. Then, $n_T^*(\theta_T)$ can be rewritten as:

$$n_T^*(\theta_T) = c_T^*(\theta_T) + \sum_{\tau=t+1}^{T-1}(1+r)^{T-\tau}\Big(c_\tau^*(\theta_\tau) - n_\tau^*(\theta_\tau)\Big) - s_t(1+r)^{T-t}$$

Substituting into S_t, we have:

$$S_t = E_{\theta_t}\Big[\theta_t u(c_t(\theta_t)) - n_t(\theta_t)\Big] + s_t(1+r)^{T-t} + V_{t+1}$$

with:

$$V_{t+1} = \sum_{\tau=t+1}^{T} E_{\theta_\tau}\Big[\theta_\tau u(c_\tau^*(\theta_\tau)) - c_\tau^*(\theta_\tau)(1+r)^{T-\tau}\Big] + \sum_{\tau=t+1}^{T-1} E_{\theta_\tau}\Big[n_\tau^*(\theta_\tau)\Big((1+r)^{T-\tau}-1\Big)\Big]$$

Since S_t is increasing in s_t, then $(\mathrm{B})_{\mathrm{t}}$ is binding. Thus, we have proved that $(\mathrm{B})_{\mathrm{T}-1}$ is binding and that $(\mathrm{B})_{\mathrm{t}}$ is binding if $(\mathrm{B})_{\mathrm{t}+1}$ to $(\mathrm{B})_{\mathrm{T}-1}$ are binding. The combination of both results implies that $(\mathrm{B})_{\mathrm{t}}$ is binding for all t. In words, it is optimal not to waste resources.

Incentive compatibility and labor constraint. Given that $n_t(\theta_t) = \theta_t\, u(c_t(\theta_t))-U_t(\theta_t)$ and that $(\mathrm{B})_{\mathrm{t}}$ is binding, the objective function of the principal at date t can be rewritten as:

$$S_t = E_{\theta_t}\Big[(1+r)^{T-t}\Big(\theta_t u(c_t(\theta_t))-c_t(\theta_t)\Big)+U_t(\theta_t)\Big(1-(1+r)^{T-t}\Big)\Big]+(1+r)^{T-t+1}s_{t-1}+V_{t+1}$$

which is decreasing in $U_t(\theta_t)$. Note also that, provided $(\mathrm{IC}_2)_{\mathrm{t}}$ is satisfied, then:

$$\dot{B}^l(\theta_t) = \dot{B}^u(\theta_t) = u(c_t(\theta_t)) + \theta_t\, u'(c_t(\theta_t))\dot{c}_t(\theta_t)) \geqslant \dot{U}_t(\theta_t) = u(c_t(\theta_t)) > 0$$

In words, the slope of the equilibrium utility is positive but smaller than the slopes of the labor feasibility constraints $B^l(\theta_t)$ and $B^u(\theta_t)$. Since we just proved that the objective function is decreasing in $U_t(\theta_t)$ (rents must be minimized), it means that $(\mathrm{IC}_1)_t$ binds at the top, that is, $U_t(\theta_t)$ binds on $B^l(\theta_t)$ at $\theta_t = \overline{\theta}$ (this, in turn, implies that $n_t(\overline{\theta}) = \bar{n}$). Let us assume that $(\mathrm{IC}_1)_t$ does not bind at any other point. Given the previous inequalities, this is true if $U_t(\underline{\theta}) < B^u(\underline{\theta})$ or, equivalently, if $n_t(\underline{\theta}) > 0$. We will neglect this inequality and check it ex-post. We then have:

$$U_t(\theta_t) = -\int_{\theta_t}^{\overline{\theta}} u(c_t(s))ds + B^l(\overline{\theta})$$

Optimal consumption. Combining the previous findings and using the standard integration by parts technique, we have:

$$S_t = E_{\theta_t}\left[(1+r)^{T-t}\Big(\theta_t\, u(c_t(\theta_t)) - c_t(\theta_t)\Big) - \Big(1-(1+r)^{T-t}\Big)u(c_t(\theta_t))\frac{F(\theta_t)}{f(\theta_t)}\right]$$
$$+\Big(\overline{\theta}\, u(c_t(\overline{\theta})) - \bar{n}\Big)\Big(1-(1+r)^{T-t}\Big) + (1+r)^{T-t+1}s_{t-1} + V_{t+1}$$

So the optimal consumption maximizes S_t under $(\mathrm{IC}_2)_\mathrm{t}$ and $(\mathrm{FC})_\mathrm{t}$. Denote by $\hat{c}_t(\theta_t)$ the consumption level that maximizes the first part of the equation:

$$u'(\hat{c}_t(\theta_t))\left[(1+r)^{T-t}\,\theta_t - \big(1-(1+r)^{T-t}\big)\,\frac{F(\theta_t)}{f(\theta_t)}\right] = (1+r)^{T-t}$$

Differentiating this expression it comes that $\hat{c}_t(\theta_t)$ is increasing in θ_t. Thus, in the absence of the term $c_t(\overline{\theta})$ in S_t, $\hat{c}_t(\theta_t)$ would be the optimal consumption. Note however that by setting a consumption $\hat{c}_t(\overline{\theta})$ for an agent with valuation $\overline{\theta}$, the principal is giving rents $\overline{\theta}u(\hat{c}_t(\overline{\theta}))$ to all the agents below that valuation. In order to decrease these rents, the principal *might* prefer to constrain consumption above a certain cutoff.[22] Overall, the solution that maximizes S_t and satisfies $(\mathrm{IC}_2)_t$ has a cutoff consumption a_t such that:

$$c_t^*(\theta_t) = \begin{cases} \hat{c}_t(\theta_t) & \forall\, \theta < \theta_t^*(a_t) \\ a_t & \forall\, \theta \geqslant \theta_t^*(a_t) \end{cases}$$

where $\hat{c}_t(\theta_t^*(a_t)) = a_t$. The only remaining issue is to determine the value a_t. Three cases are possible: $a_t > \overline{a}_t \equiv \hat{c}_t(\overline{\theta})$; $a_t < \underline{a}_t \equiv \hat{c}_t(\underline{\theta})$; $a_t \in [\underline{a}_t, \overline{a}_t]$. Let:

$$\Psi_t(\theta_t, x) = \left[(1+r)^{T-t}\Big(\theta_t\, u(x) - x\Big) - \Big(1-(1+r)^{T-t}\Big)u(x)\frac{F(\theta_t)}{f(\theta_t)}\right]$$

For all $a_t > \overline{a}_t$, the welfare is:

$$\int_{\underline{\theta}}^{\overline{\theta}} \Psi_t(\theta_t, \hat{c}_t(\theta_t))\, dF(\theta_t) + \Big(\overline{\theta}\, u(a_t) - \bar{n}\Big)\Big(1-(1+r)^{T-t}\Big) + (1+r)^{T-t+1}s_{t-1} + V_{t+1}$$

This function is decreasing in a_t, so the principal always chooses $a_t \leqslant \overline{a}_t$. For all $a_t \in [\underline{a}_t, \overline{a}_t]$, the welfare of the principal in equilibrium is:

$$S_t(a_t) = \int_{\underline{\theta}}^{\theta_t^*(a_t)} \Psi_t(\theta_t, \hat{c}_t(\theta_t))dF(\theta_t) + \int_{\theta_t^*(a_t)}^{\overline{\theta}} \Psi_t(\theta_t, a_t)dF(\theta_t)$$

[22] This is a technical difference of our analysis relative to standard programs. Typically, the utility at the endpoint (where the individual rationality (IR) constraint binds) is exogenous. In our setting (with no IR constraint) the utility at the endpoint $U_t(\overline{\theta})$ is mechanism dependent, that is, it is affected by $c(\overline{\theta})$.

$$+\Big(\overline{\theta}\,u(a_t)-\bar{n}\Big)\Big(1-(1+r)^{T-t}\Big)+(1+r)^{T-t+1}s_{t-1}+V_{t+1}$$

The optimal consumption cap a_t is the one that maximizes $S_t(a_t)$. We have:

$$S_t'(a_t)=u'(a_t)K_t(a_t)-(1+r)^{T-t}\Big(1-F(\theta_t^*(a_t))\Big)\qquad\text{and}\qquad S_t''(a_t)=u''(a_t)K_t(a_t)$$

where

$$K_t(a_t)=\int_{\theta^*(a_t)}^{\overline{\theta}}\left[(1+r)^{T-t}\,\theta_t-\Big(1-(1+r)^{T-t}\Big)\frac{F(\theta_t)}{f(\theta_t)}\right]f(\theta_t)d\theta_t+\Big(1-(1+r)^{T-t}\Big)\overline{\theta}$$

Note that $K_t(a_t)$ is decreasing in a_t and that $K_t(\overline{a}_t)<0$ therefore $\overline{a}_t$ is never optimal (there is always bunching at the top). We have two cases.

If $K_t(\underline{a}_t)<0$, then $S_t'(a_t)<0$ for all $a_t\in[\underline{a}_t,\overline{a}_t]$. The optimal consumption level a_t is in $[0,\underline{a}_t]$. Therefore $c_t^*(\theta_t)=a_t$ for all $\theta_t\in[\underline{\theta},\overline{\theta}]$ and a_t maximizes:

$$\int_{\underline{\theta}}^{\overline{\theta}}\Psi_t(\theta_t,a_t)\,dF(\theta_t)+\Big(\overline{\theta}\,u(a_t)-\bar{n}\Big)\Big(1-(1+r)^{T-t}\Big)+(1+r)^{T-t+1}s_{t-1}+V_{t+1}$$

If $K_t(\underline{a}_t)>0$, there exists $\hat{a}_t\in(\underline{a}_t,\overline{a}_t)$ such that $K_t(\hat{a}_t)=0$. The welfare is strictly decreasing when $a_t\geqslant\hat{a}_t$ and it is concave when $a_t\in(\underline{a}_t,\hat{a}_t)$. If $S_t'(\underline{a}_t)<0$, we are in the same case as before (bunching for all θ_t). Last, if $S_t'(\underline{a}_t)>0$, then there exists an interior maximum $a_t^*\in(\underline{a}_t,\hat{a}_t)$ and the cutoff valuation is $\theta_t^*\equiv\theta_t^*(a_t^*)$.

Optimal labor. Given that $n_t^*(\theta_t)=\theta_t\,u(c_t^*(\theta_t))-U_t(\theta_t)$, we have:

$$n_t^*(\theta_t)=\bar{n}-\left[\overline{\theta}\,u(c_t^*(\overline{\theta}))-\theta_t\,u(c_t^*(\theta_t))-\int_{\theta_t}^{\overline{\theta}}u(c_t^*(s))ds\right]$$

In particular, for all $\theta_t\geqslant\theta_t^*(a_t^*)$, there is bunching and $n_t^*(\theta_t)=\bar{n}$. Also,

$$\frac{dn_t^*}{d\theta_t}=\theta_t\,u'(c_t^*(\theta_t))\,\frac{dc_t^*}{d\theta_t}$$

which is strictly positive for all $\theta_t<\theta_t^*$. Last, the neglected inequality $n_t^*(\underline{\theta})>0$ is automatically satisfied if $\bar{n}$ is "sufficiently large" or, more specifically, if:

$$\bar{n}>\overline{\theta}\,u(c_t^*(\overline{\theta}))-\underline{\theta}\,u(c_t^*(\underline{\theta}))-\int_{\underline{\theta}}^{\overline{\theta}}u(c_t^*(s))ds$$

Appendix B. Proof of Propositions 4 and 5. Let $W(\theta) = \theta\, w(x(\theta)) + v(y(\theta))$. Using standard techniques (proof omitted), the incentive compatibility constraints ($\hat{\text{IC}}$) in program $\mathcal{U}_{\text{AI}}$ are equivalent to the following first- and second-order conditions:

$$\dot{W}(\theta) = w(x(\theta)) \qquad \text{and} \qquad \dot{x}(\theta) \geqslant 0$$

Also, when $v' > 0$ and $r' < 0$ or when $v' < 0$ and $r' > 0$, ($\hat{\text{BB}}$) can be rewritten as:

$$W(\theta) \leqslant \theta\, w(x(\theta)) + v(r^{-1}(x(\theta)))$$

Since $v(y(\theta)) = W(\theta) - \theta w(x(\theta))$, program $\mathcal{U}_{\text{AI}}$ can thus be rewritten as:

$$\begin{aligned}
\mathcal{U}_{\text{AI}}: \max_{\{(x(\theta),W(\theta))\}} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \Big[\theta\, u(x(\theta)) - \theta\, w(x(\theta)) + W(\theta)\Big]\, dF(\theta) \\
\text{s.t.} \quad & \dot{W}(\theta) = w(x(\theta)) && (\hat{\text{IC}}_1) \\
& \dot{x}(\theta) \geqslant 0 && (\hat{\text{IC}}_2) \\
& W(\theta) \leqslant B(\theta) = \theta\, w(x(\theta)) + v(r^{-1}(x(\theta))) && (\hat{\text{BB}})
\end{aligned}$$

The equilibrium utility increases at rate $\dot{W}(\theta) = w(x(\theta))$ and the upper bound of ($\hat{\text{BB}}$) increases at rate $\dot{B}(\theta) \equiv \dot{x}(\theta)\Big[\theta w'(x(\theta)) + v'(r^{-1}(x(\theta)))r^{-1'}(x(\theta))\Big] + w(x(\theta))$. Given ($\hat{\text{IC}}_2$), assumption 1 and the definition of $x^D(\theta)$ as the maximum in $\mathcal{W}$, then in equilibrium:

$$\dot{W}(\theta) \lessgtr \dot{B}(\theta) \;\Leftrightarrow\; x(\theta) \lessgtr x^D(\theta).$$

Since $x^F(\theta) < x^D(\theta)$, then $(x(\theta'), y(\theta'))$ with $x(\theta') > x^D(\theta')$, yields lower utility to the principal than $(x^D(\theta'), r^{-1}(x^D(\theta')))$, provided the latter is incentive compatible at θ'. The indifference curves of the principal satisfy $x'(y) = -v'(y)/\theta u'(x)$. They are decreasing and convex if $v' > 0$ and $r' < 0$ and increasing and convex if $v' < 0$ and $r' > 0$. To satisfy incentive compatibility, $dx/dy = -v'(y)/\theta w'(x)$. Assume now that the contract entails $(x^D(\theta'), y^{ic}(\theta'))$ for some θ' with $x^D(\theta') < r(y^{ic}(\theta'))$. Consider a deviation to $x(\theta') > x^D(\theta')$ and let $y(\theta')$ be such that $(x(\theta'), y(\theta'))$ is incentive compatible. Given the previous properties, $\theta\, u(x^D(\theta')) + v(y^{ic}(\theta')) > \theta\, u(x(\theta')) + v(y(\theta'))$. This proves that it is never optimal to set $x(\theta) > x^D(\theta)$ for any θ. Therefore, from now on, we shall restrict the attention to solutions of the form $x(\theta) \leqslant x^D(\theta)$ for all θ.

Note that $W(\theta)$ enters positively in the principal's objective function. Also, $x(\theta) \leqslant x^D(\theta)$ implies $\dot{W}(\theta) \leqslant \dot{B}(\theta)$. Combining both arguments, $W(\theta)$ binds in ($\hat{\text{BB}}$) at the lower bound $\underline{\theta}$. Using ($\hat{\text{IC}}_1$) and ($\hat{\text{BB}}$), we then have:

$$W(\theta) = \int_{\underline{\theta}}^{\theta} w(x(s))ds + W(\underline{\theta}) \quad \text{with} \quad W(\underline{\theta}) = \underline{\theta} w(x(\underline{\theta})) + v(r^{-1}(x(\underline{\theta})))$$

Using standard integration by parts techniques, the problem becomes:

$$\begin{aligned}
\mathcal{U}_{\text{AI}}: \quad \max_{\{x(\theta)\}} \quad & \int_{\underline{\theta}}^{\overline{\theta}} \left[\theta\, u(x(\theta)) - \theta\, w(x(\theta)) + w(x(\theta))\frac{1-F(\theta)}{f(\theta)}\right] dF(\theta) + W(\underline{\theta}) \\
\text{s.t.} \quad & \dot{x}(\theta) \geqslant 0 && (\hat{\text{IC}}_2)\\
& W(\underline{\theta}) = \underline{\theta} w(x(\underline{\theta})) + v(r^{-1}(x(\underline{\theta}))) && (\text{E})\\
& x(\theta) \leqslant x^D(\theta) && (\text{D})
\end{aligned}$$

where (E) is the utility at $\underline{\theta}$ and (D) is the restriction on the domain. The rest of the proof proceeds as follows. First, we ignore ($\hat{\text{IC}}_2$) and (E) and find the solutions that satisfy (D). Second, we construct the solutions that also satisfy ($\hat{\text{IC}}_2$). Last, we introduce (E). Let:

$$\Lambda(x,\theta) = \theta\, T(w(x)) - \theta\, w(x) + w(x)\frac{1-F(\theta)}{f(\theta)}.$$

where $\Lambda(0,\theta) = 0$; $\frac{\partial \Lambda(x,\theta)}{\partial x} = w'(x)\Big[\theta T'(w(x)) - \theta + \frac{1-F(\theta)}{f(\theta)}\Big]$; $\frac{\partial \Lambda(x,\theta)}{\partial x}\Big|_{\overline{\theta}} \leqslant 0$; $\frac{\partial^2 \Lambda(x,\theta)}{\partial x \partial \theta} = w'(x)\Big[T'(w(x)) - 1 + \Big(\frac{1-F(\theta)}{f(\theta)}\Big)'\Big] \leqslant 0$; $\frac{\partial^2 \Lambda(x,\theta)}{\partial x^2} = \frac{w''(x)}{w'(x)}\frac{\partial \Lambda(x,\theta)}{\partial x} + [w'(x)]^2\, \theta\, T''(w(x))$. Denote by $\tilde{x}(\theta)$ the interior optimum of $\Lambda(x,\theta)$, if it exists. We shall consider two different cases.

Case 1: $T''(\cdot) > 0$. $\frac{\partial \Lambda(\tilde{x}(\theta),\theta)}{\partial x} = 0$ implies $\frac{\partial^2 \Lambda(\tilde{x}(\theta),\theta)}{\partial x^2} > 0$, so $\tilde{x}(\theta)$ is the unique minimum of $\Lambda(x,\theta)$. The maxima are the corner solutions 0 or $x^D(\theta)$. Also, there exists $\tilde{\theta}$ such that for all $\theta > \tilde{\theta}$, $\Lambda(x,\theta)$ is strictly decreasing in x and the maximum is 0. For $\theta \leqslant \tilde{\theta}$, the maximum alternates between 0 and $x^D(\theta)$.

Case 1a. Suppose that the maximum at $\underline{\theta}$ is $x^D(\underline{\theta})$. Then, there exists a series of cutoffs $(\theta_0, \ldots, \theta_{2t-1}, \theta_{2t})$ where $\theta_0 = \underline{\theta}$, $\theta_{2t-1} = \tilde{\overline{\theta}}$ and $\theta_{2t} = \overline{\theta}$, such that:

$$\tilde{x}(\theta) = \begin{cases} x^D(\theta) & \text{if} \quad \theta \in [\theta_s, \theta_{s+1}] \\ 0 & \text{if} \quad \theta \in (\theta_{s+1}, \theta_{s+2}) \end{cases} \qquad \forall\ s \in \{0, 2, \ldots, 2t-2\}$$

Note that $\tilde{x}(\theta)$ does not satisfy ($\hat{\text{IC}}_2$) in the neighborhood of θ_{s+1}. When adding this constraint, we could set consumption at $x^D(\theta_{s+1})$ for all $\theta \in (\theta_{s+1}, \theta_{s+2})$ (it is obviously suboptimal to go above). It may however, be preferable to start pooling at $\theta'_{s+1} < \theta_{s+1}$: the cost of $x^D(\theta'_{s+1}) < x^D(\theta)\ \forall\, \theta \in (\theta'_{s+1}, \theta_{s+1}]$ may be offset by the benefits of $x^D(\theta'_{s+1}) < x^D(\theta_{s+1})\ \forall\, \theta \in (\theta_{s+1}, \theta_{s+2})$.[23] Overall, there will exist new cutoffs $\theta'_{s+1} \in [\theta_s, \theta_{s+1}]$ such that the solution that maximizes the principal's objective under (D) and ($\hat{\text{IC}}_2$) is:

$$x^*(\theta) = \begin{cases} x^D(\theta) & \text{if} \quad \theta \in [\theta_s, \theta'_{s+1}] \\ x^D(\theta'_{s+1}) & \text{if} \quad \theta \in (\theta'_{s+1}, \theta_{s+2}) \end{cases} \qquad \forall\ s \in \{0, 2, \ldots, 2t-2\}$$

[23] The argument is the same as to where bunching should start in standard mechanism design problems when $\dot{x}(\theta) \geqslant 0$ is not automatically satisfied.

Last, let a be the optimal consumption at $\underline{\theta}$, where $a \leqslant x^D(\underline{\theta})$ to satisfy (D). Denote by $\hat{x}(\theta)$ the optimal solution of the principal's program under ($\hat{\text{IC}}_2$), (E), (D). We have $\hat{x}(\underline{\theta}) = a$ and $\hat{x}(\theta) = x^*(\theta)\ \forall\, \theta > \underline{\theta}$. The equilibrium utility of the principal is then:

$$\int_{\underline{\theta}}^{\overline{\theta}} \Lambda(x^*(\theta), \theta) dF(\theta) + \underline{\theta} w(a) + v(r^{-1}(a))$$

This utility is increasing in a, so $a = x^D(\underline{\theta})$. Overall, the optimal solution is:

$$\hat{x}(\theta) = x^*(\theta) = \left\{ \begin{array}{lll} x^D(\theta) & \text{if} & \theta \in [\theta_s, \theta'_{s+1}] \\ x^D(\theta'_{s+1}) & \text{if} & \theta \in (\theta'_{s+1}, \theta_{s+2}) \end{array} \right. \qquad \forall\ s \in \{0, 2, \ldots, 2t-2\}$$

It remains to determine $\hat{y}(\theta)$. The agent's utility under delegation is:

$$\begin{aligned} W^D(\theta) &= \theta\, w(x^D(\theta)) + v(r^{-1}(x^D(\theta)) && (1) \\ &= \int_{\underline{\theta}}^{\theta} w(x^D(c))dc + \underline{\theta}\, w(x^D(\underline{\theta})) + v(r^{-1}(x^D(\underline{\theta}))) && (2) \end{aligned}$$

since $\dot{W}^D(\theta) = w(x^D(\theta))$. The agent's utility under the optimal contract $(\hat{x}(\theta), \hat{y}(\theta))$ is:

$$\begin{aligned} W(\theta) &= \theta\, w(\hat{x}(\theta)) + v(\hat{y}(\theta)) && (3) \\ &= \int_{\underline{\theta}}^{\theta} w(\hat{x}(c))dc + \underline{\theta}\, w(\hat{x}(\underline{\theta})) + v(r^{-1}(\hat{x}(\underline{\theta}))) && (4) \end{aligned}$$

For all $\theta \in [\underline{\theta}, \theta'_1]$, we have $\hat{x}(\theta) = x^D(\theta)$ and $W(\theta) = W^D(\theta)$, so $v(\hat{y}(\theta)) = v(r^{-1}(x^D(\theta)))$, and resources are not wasted. For all $\theta \in (\theta'_1, \theta_2)$, we have $\hat{x}(\theta) = x^D(\theta'_1)$. Using (2) and (4), we have $W(\theta) = W^D(\theta'_1) + (\theta - \theta'_1) w(x^D(\theta'_1))$. Using (1) and (3), we get $v(\hat{y}(\theta)) = v(r^{-1}(x^D(\theta'_1)))$ and, again, resources are not wasted. For all $\theta \in [\theta_2, \theta'_3]$, we have $\hat{x}(\theta) = x^D(\theta)$ but $W(\theta) < W^D(\theta)$. Then, $v(\hat{y}(\theta)) < v(r^{-1}(x^D(\theta)))$, that is, for all $\theta \geqslant \theta_2$ resources are wasted.

Case 1b. Suppose that the maximum at $\underline{\theta}$ is 0. Following the analogous reasoning as in case 1a, the maximization of the principal's objective under ($\hat{\text{IC}}_2$) and (D) yields:

$$x^*(\theta) = \left\{ \begin{array}{lll} 0 & \text{if} & \theta \in [\underline{\theta}, \theta_1) \\ x^D(\theta) & \text{if} & \theta \in [\theta_s, \theta'_{s+1}] \qquad \forall\ s \in \{1, 3, \ldots, 2t-1\} \\ x^D(\theta'_{s+1}) & \text{if} & \theta \in (\theta'_{s+1}, \theta_{s+2}) \quad \forall\ s \in \{1, 3, \ldots, 2t-1\} \end{array} \right.$$

Adding constraint (E) to the program, modifies the solution into $\hat{x}(\theta) = a$ for all $\theta \in [\underline{\theta}, \theta_1)$ and $\hat{x}(\theta) = x^*(\theta)$ for all $\theta \in [\theta_1, \overline{\theta}]$, with $a \in [0, x^D(\underline{\theta})]$. The principal's utility is then:

$$\int_{\underline{\theta}}^{\theta_1} \Lambda(a, \theta)\, dF(\theta) + \int_{\theta_1}^{\overline{\theta}} \Lambda(x^*(\theta), \theta)\, dF(\theta) + \underline{\theta}\, w(a) + v(r^{-1}(a))$$

Let $\hat{a} = \text{argmax}_{a\in[0,x^D(\underline{\theta})]} \int_{\underline{\theta}}^{\theta_1} \Lambda(a,\theta)dF(\theta) + \underline{\theta}\, w(a) + v(r^{-1}(a))$. The optimal solution is:

$$\hat{x}(\theta) = \begin{cases} \hat{a} & \text{if} \quad \theta \in [\underline{\theta}, \theta_1) \\ x^D(\theta) & \text{if} \quad \theta \in [\theta_s, \theta'_{s+1}] \quad \forall\ s \in \{1,3,\ldots,2t-1\} \\ x^D(\theta'_{s+1}) & \text{if} \quad \theta \in (\theta'_{s+1}, \theta_{s+2}) \quad \forall\ s \in \{1,3,\ldots,2t-1\} \end{cases}$$

Using the same method as in case 1a, we can compute $\hat{y}(\theta)$. For all $\theta < \theta_1$, $\hat{x}(\theta) = \hat{a} \leqslant x^D(\theta)$ and $W(\theta) = \int_{\underline{\theta}}^{\theta} w(\hat{a})ds + \underline{\theta}\, w(\hat{a}) + v(r^{-1}(\hat{a}))$. Combining it with (3), we get that $v(\hat{y}(\theta)) = v(r^{-1}(\hat{a}))$, so resources are not wasted. For all $\theta \in [\theta_1, \theta'_2]$, consumption is $x^D(\theta)$ and, using (2) and (4), we have $W(\theta) < W^D(\theta)$. Therefore, $v(\hat{y}(\theta)) < v(r^{-1}(x^D(\theta)))$ and resources are wasted for all $\theta \geqslant \theta_1$.

Case 2: $T''(\cdot) \leqslant 0$. If $\tilde{x}(\theta)$ exists, it is the unique interior maximum. However, it is decreasing in θ so it does not satisfy ($\hat{\text{IC}}_2$). Again, there exists $\tilde{\theta}$ such that for all $\theta > \tilde{\theta}$, $\Lambda(x,\theta)$ is strictly decreasing in x, so the maximum is 0. For all $\theta \leqslant \tilde{\theta}$, $\tilde{x}(\theta)$ exists. The maximum of $\Lambda(x,\theta)$ under (D), is $\tilde{x}(\theta)$ if $\tilde{x}(\theta) \leqslant x^D(\theta)$ and $x^D(\theta)$ if $\tilde{x}(\theta) \geqslant x^D(\theta)$.

Case 2a. Since $\frac{dx^D(\theta)}{d\theta} > 0$ and $\frac{d\tilde{x}(\theta)}{d\theta} < 0$, if $x^D(\underline{\theta}) \leqslant \tilde{x}(\underline{\theta})$, then there exists θ' such that $x^D(\theta) < \tilde{x}(\theta)$ for all $\theta < \theta'$ and $x^D(\theta) \geqslant \tilde{x}(\theta)$ for all $\theta \geqslant \theta'$. To satisfy ($\hat{\text{IC}}_2$), the principal could set $x^D(\theta)$ for all $\theta < \theta'$ and $x^D(\theta')$ for all $\theta \geqslant \theta'$. However, using the same logic as in case 1a, there will exist a cutoff $\hat{\theta} \in [\underline{\theta}, \theta')$ such that (see later for its determination):

$$x^*(\theta) = \begin{cases} x^D(\theta) & \text{if} \quad \theta \in [\underline{\theta}, \hat{\theta}) \\ x^D(\hat{\theta}) & \text{if} \quad \theta \in [\hat{\theta}, \overline{\theta}] \end{cases}$$

Adding constraint (E) to the program modifies the solution into $\hat{x}(\underline{\theta}) = a$ and $\hat{x}(\theta) = x^*(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta}]$, with $a \in [0, x^D(\underline{\theta})]$. The principal's equilibrium utility is then:

$$\int_{\underline{\theta}}^{\overline{\theta}} \Lambda(x^*(\theta),\theta)dF(\theta) + \underline{\theta}\, w(a) + v(r^{-1}(a))$$

which is increasing in a, so $a = x^D(\underline{\theta})$. Overall, the optimal solution is:

$$\hat{x}(\theta) = \begin{cases} x^D(\theta) & \text{if} \quad \theta \in [\underline{\theta}, \hat{\theta}) \\ x^D(\hat{\theta}) & \text{if} \quad \theta \in [\hat{\theta}, \overline{\theta}] \end{cases}$$

Using the same reasoning as in case 1a, resources are never wasted. Last and for the sake of completeness, we characterize $\hat{\theta}$. Given $\hat{\theta}$, the equilibrium utility of the principal is:

$$\hat{U} = \int_{\underline{\theta}}^{\hat{\theta}} \Lambda(x^D(\theta),\theta)dF(\theta) + \int_{\hat{\theta}}^{\overline{\theta}} \Lambda(x^D(\hat{\theta}),\theta)dF(\theta) + \underline{\theta}\, w(x^D(\underline{\theta})) + v(r^{-1}(x^D(\underline{\theta})))$$

The first-order condition that determines the optimal cutoff $\hat{\theta}$ is then given by (note that we would need to impose further restrictions to ensure uniqueness):

$$\frac{d\hat{U}}{d\hat{\theta}} = 0 \;\Rightarrow\; \int_{\hat{\theta}}^{\overline{\theta}} \frac{\partial \Lambda}{\partial x}(x^D(\hat{\theta}), \theta)\, dF(\theta) = 0$$

Since $\left.\dfrac{d\hat{U}}{d\hat{\theta}}\right|_{\hat{\theta}=\theta'} = \displaystyle\int_{\theta'}^{\overline{\theta}} \frac{\partial \Lambda(x^D(\theta'), \theta)}{\partial x} \frac{\partial x^D(\theta')}{\partial \theta}\, dF(\theta) < 0$, we then have that $\hat{\theta} < \theta'$.

Case 2b. Since $\frac{d\tilde{x}(\theta)}{d\theta} < 0$, if $x^D(\underline{\theta}) > \tilde{x}(\underline{\theta})$, then it is optimal to set the same consumption level for all θ. This amount is given by:

$$\hat{x}(\theta) = \hat{a} \quad \forall\, \theta \in [\underline{\theta}, \overline{\theta}] \quad \text{where} \;\; \hat{a} = \arg\max_a \int_{\underline{\theta}}^{\overline{\theta}} \Lambda(a, \theta) dF(\theta) + \underline{\theta}\, w(a) + v(r^{-1}(a))$$

Note that $\frac{\partial \Lambda}{\partial a}(x^D(\underline{\theta}), \theta) < 0$ and $\frac{d}{da}\Big[\underline{\theta}\, w(x^D(\underline{\theta})) + v(r^{-1}(x^D(\underline{\theta})))\Big] = 0$, so $\hat{a} < x^D(\underline{\theta})$.

Case 3: Special case $T''(\cdot) = 0$. Assume $w(x) = \alpha u(x)$ with $\alpha > 1$. We have:

$$\Lambda(x, \theta) = w(x)\, K(\theta) \qquad \text{where} \;\; K(\theta) = \theta \frac{1}{\alpha} - \theta + \frac{1 - F(\theta)}{f(\theta)}$$

Following the same reasoning as in case 2a, we have $\hat{x}(\theta) = x^D(\theta)$ if $\theta \in [\underline{\theta}, \theta_l)$ and $\hat{x}(\theta) = x^D(\theta_l)$ if $\theta \in [\theta_l, \overline{\theta}]$, where the cutoff θ_l is determined by the following equality:

$$\frac{d\hat{U}}{d\theta_l} = 0 \Rightarrow w'(x^D(\theta_l)) \frac{dx^D(\theta_l)}{d\theta} \int_{\theta_l}^{\overline{\theta}} K(\theta) f(\theta)\, d\theta = 0 \Rightarrow \int_{\theta_l}^{\overline{\theta}} \left[(\frac{1}{\alpha} - 1)\theta f(\theta) + 1 - F(\theta) \right] d\theta = 0 \quad (5)$$

$$\Rightarrow E[\theta \mid \theta > \theta_l] \;=\; \alpha\, \theta_l \quad (6)$$

Note that the cutoff θ_l is indeed a unique maximum:

$$\frac{d^2\hat{U}}{d\theta_l^2} = \frac{d}{d\theta_l}\left[w'(x^D(\theta_l)) \frac{dx^D(\theta_l)}{d\theta} \right] \int_{\theta_l}^{\overline{\theta}} K(\theta) f(\theta)\, d\theta - w'(x^D(\theta_l)) \frac{dx^D(\theta_l)}{d\theta} K(\theta_l) f(\theta_l) < 0$$

where the first term is equal to zero by (5) and $K(\theta_l) > 0$ by (5) and $dK/d\theta < 0$. Also, every type consumes the same amount ($\theta_l = \underline{\theta}$) if and only if $\left.\frac{d\hat{U}}{d\theta_l}\right|_{\theta_l = \underline{\theta}} \leqslant 0 \;\Rightarrow \alpha > \overline{\alpha} \equiv \frac{E[\theta]}{\underline{\theta}}$.

Differentiating (5): $-K(\theta_l(\alpha), \alpha) f(\theta_l(\alpha)) \dfrac{d\theta_l}{d\alpha} + \displaystyle\int_{\theta_l}^{\overline{\theta}} \frac{\partial K(\theta, \alpha)}{\partial \alpha} f(\theta) d\theta = 0 \;\Rightarrow\; \dfrac{d\theta_l}{d\alpha} < 0.$

Last, if $F(\theta) > G(\theta)$ for all $\theta \in (\underline{\theta}, \overline{\theta})$, then $E_{G(\theta)}[\theta \mid \theta > \tilde{\theta}] > E_{F(\theta)}[\theta \mid \theta > \tilde{\theta}]$ for all $\tilde{\theta}$. As a result and given (6), $\theta_l^G > \theta_l^F$ where θ_l^G is the cutoff under distribution $G(\theta)$ and θ_l^F is the cutoff under distribution $F(\theta)$.

References

1. Ainslie, George. 1992. *Picoeconomics*. Cambridge: Cambridge University Press.

2. Amador, Manuel, Ivan Werning, and George-Marios Angeletos. 2006. "Commitment vs. Flexibility." *Econometrica*, 74(2): 365-96.

3. Ariely, Dan, George Loewenstein, and Drazen Prelec. 2003. "Coherent Arbitrariness: Stable Demand Curves Without Stable Preferences." *Quarterly Journal of Economics*, 118(1): 73-105.

4. Ariely, Dan, George Loewenstein, and Drazen Prelec. 2006. "Tom Sawyer and the Construction of Value." *Journal of Economic Behavior & Organization*, 60(1): 1-10.

5. Bataglini, Marco, Roland Bénabou, and Jean Tirole. 2005. "Self-Control in Peer Groups." *Journal of Economic Theory*, 112(4): 848-87.

6. Baumeister, Roy. 2003. "The Psychology of Irrationality: Why People Make Foolish, Self-Defeating Choices." In *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, ed. Isabelle Brocas and Juan D. Carrillo, 3-16. Oxford: Oxford University Press.

7. Becerra, Lino, Hans Breiter, Roy Wise, Gilberto Gonzalez, and David Borsook. 2001. "Reward Circuitry Activation by Noxious Thermal Stimuli." *Neuron*, 32(5): 927-46.

8. Bechara, Antoine. 2005. "Decision Making, Impulse Control and Loss of Willpower to Resist Drugs: a Neurocognitive Perspective." *Nature Neuroscience*, 8(11): 1458-1463.

9. Bechara, Antoine, Hanna Damasio, Antonio Damasio, and Gregory Lee. 1999. "Different Contributions of the Human Amygdala and Ventromedial Prefrontal Cortex to Decision-Making." *Journal of Neuroscience*, 19(13): 5473-81.

10. Bechara, Antoine, Daniel Tranel, Hanna Damasio, Ralph Adolphs, Charles Rockland, and Antonio Damasio. 1995. "Double Dissociation of Conditioning and Declarative Knowledge Relative to the Amygdala and Hippocampus in Humans." *Science*, 269: 1115-1118.

11. Bem, Daryl. 1967. "Self-Perception: an Alternative Interpretation of Cognitive Dissonance Phenomena." *Psychological Review*, 74(3): 183-200.

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

13. Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy*, 112(4): 848-87.

14. Benhabib, Jess, and Alberto Bisin. 2005. "Modeling Internal Commitment Mechanisms and Self-Control: a Neuroeconomics Approach to Consumption-Saving Decisions." *Games and Economic Behavior*, 52(2): 460-92.

15. Bernheim, B. Douglas, and Antonio Rangel. 2004. "Addiction and Cue-Triggered Decision Processes." *American Economic Review*, 94(5): 1558-90.

16. Berns, Gregory, Jonathan Cohen, and Mark Mintun. 1997. "Brain Regions Responsive to Novelty in the Absence of Awareness." *Science*, 276: 1272-75.

17. Berns, Gregory, Samuel McClure, Giuseppe Pagnoni, and Read Montague. 2001. "Predictability Modulates Human Brain Response to Reward." *Journal of Neuroscience*, 21(8): 2793-98.

18. Berridge, Kent. 2003. "Irrational Pursuit: Hyper-Incentives from a Visceral Brain." In *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, ed. Isabelle Brocas and Juan D. Carrillo, 17-40. Oxford: Oxford University Press.

19. Berridge, Kent, and Terry Robinson. 2003. "Parsing Reward." *Trends in Neurosciences*, 26(9): 507-13.

20. Bodner, Ronit, and Drazen Prelec. 2003. "Self-Signaling and Diagnostic Utility in Everyday Decision Making." In *The Psychology of Economic Decisions. Vol.1: Rationality and Well-Being*, ed. Isabelle Brocas and Juan D. Carrillo, 105-126. Oxford: Oxford University Press.

21. Brocas, Isabelle, and Juan D. Carrillo. 2004. "Entrepreneurial Boldness and Excessive Investment." *Journal of Economics & Management Strategy*, 13(2): 321-50.

22. Brocas, Isabelle, and Juan D. Carrillo. Forthcoming. "Theories of the Mind." *American Economic Review Papers and Proceedings*.

23. Caillaud, Bernard, and Bruno Jullien. 2000. "Modelling Time-Inconsistent Preferences." *European Economic Review Papers and Proceedings*, 44(4-6): 1116-24.

24. Caillaud, Bernard, Daniel Cohen, and Bruno Jullien. 1999. "Towards a Theory of Self-Restraint." Unpublished.

25. Camerer, Colin, Linda Babcock, George Loewenstein, Richard Thaler. 1997. "Labor Supply of New York City Cabdrivers: One Day at a Time." *Quarterly Journal of Economics*, 112(2): 407-441.

26. Camerer, Colin, George Loewenstein, and Drazen Prelec. 2004. "Neuroeconomics: Why Economics Needs Brains." *Scandinavian Journal of Economics*, 106(3): 555-79.

27. Camerer, Colin, George Loewenstein, and Drazen Prelec. 2005. "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature*, 43(1): 9-64.

28. Caplin, Andrew, and John Leahy. 2001. "Psychological Expected Utility Theory and Anticipatory Feelings." *Quarterly Journal of Economics*, 116(1): 55-80.

29. Carrillo, Juan D. 2005. "To Be Consumed with Moderation." *European Economic Review*, 49(1): 99-111.

30. Carrillo, Juan D., and Thomas Mariotti. 2000. "Strategic Ignorance as a Self-Disciplining Device." *Review of Economic Studies*, 67(3): 529-544.

31. Damasio, Antonio. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: G.P. Putnam.

32. Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford University Press.

33. Delgado, Mauricio, Leigh Nystrom, Kate Fissell, Douglas Noll, and Julie Fiez. 2000. "Tracking the Hemodynamic Responses to Reward and Punishment in the Striatum." *Journal of Neurophysiology*, 84(6): 3072-77.

34. Elliott, Rebecca, Jana Newman, Olivia Longe, and William Deakin. 2003. "Differential Response Patterns in the Striatum and Orbitofrontal Cortex to Financial Reward in Humans: a Parametric fMRI study." *Journal of Neuroscience*, 23(1): 303-7.

35. Elster, Jon. 2004. "Costs and Constraints in the Economy of the Mind." In *The Psychology of Economic Decisions. Vol.2: Reasons and Choices*, ed. Isabelle Brocas and Juan D. Carrillo, 3-14. Oxford: Oxford University Press.

36. Festinger, Leon. 1957. *A Theory of Cognitive Dissonance*. Stanford: Stanford U. Press.

37. Frederick, Shane, George Loewenstein, and Edward O'Donoghue. 2002. "Time Discounting and Time Preference: A Critical Review." *Journal of Economic Literature*, 40(2): 351-401.

38. Fudenberg, Drew, and David K. Levine. 2006. "A Dual Self Model of Impulse Control." *American Economic Review*, 96(5): 1449-1476.

39. Fudenberg, Drew, and Jean Tirole. 1991. *Game Theory*. Cambridge: MIT Press.

40. Gehring, William, Brian Goss, Michael Coles, David Meyer, and Emanuel Donchin. 1993. "A Neural System for Error Detection and Compensation." *Psychological Science*, 4(6): 385-90.

41. Glimcher, Paul, Joseph Kable, and Kenway Louie. 2007. "Neuroeconomic Studies of Impulsivity: Now or Just as Soon as Possible?" *American Economic Review Papers and Proceedings*, 97(2): 142-7.

42. Gul, Faruk, and Wolfgang Pesendorfer. 2001. "Temptation and Self-Control." *Econometrica*, 69(6): 1403-35.

43. Gur, Ruben and Harold Sackeim. 1979. "Self-deception: A Concept in Search of a Phenomenon." *Journal of Personality and Social Psychology*, 37(2): 147-169.

44. Hall, Robert, and Frederick Mishkin. 1982. "The Sensitivity of Consumption to Transitory Income: Estimates from Panel Data on Households." *Econometrica*, 50(2): 461-482.

45. Heath, Chip, and Jack Soll. 1996. "Mental Budgeting and Consumer Decisions." *Journal of Consumer Research*, 23(1): 40-52.

46. Jensen, James, Anthony McIntosh, Adrian Crawley, David Mikulis, Gary Remington, and Shitij Kapur. 2003. "Direct Activation of the Ventral Striatum in Anticipation of Aversive Stimuli." *Neuron*, 40(6): 1251-57.

47. Kerns, John, Jonathan Cohen, Angus MacDonald, Raymond Cho, Andrew Stenger, and Cameron Carter. 2004. "Anterior Cingulate Conflict Monitoring and Adjustments in Control." *Science*, 303(5660): 1023-1026.

48. Knowlton, Barbara, Jennifer Mangels, and Larry Squirre. 1996. "A Neostriatal Habit Learning System in Humans." *Science*, 273(5280): 1399-1402.

49. Knutson, Brian, Andrew Westdorp, Erica Kaiser, and Daniel Hommer. 2000. "fMRI Visualization of Brain Activity during a Monetary Incentive Delay Task." *NeuroImage*, 12(1): 20-27.

50. Koopmans, Tjalling. 1960. "Stationary Ordinal Utility and Impatience." *Econometrica*, 28(2): 287-309.

51. Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *Quarterly Journal of Economics*, 112(2): 443-477.

52. LeDoux, Joseph. 1996. *The Emotional Brain. The Mysterious Underpinnings of Emotional Life*. Simon and Schuster: New York.

53. Lichtenstein, Sarah, and Paul Slovic. 2006. *The Construction of Preference*. Cambridge: Cambridge University Press.

54. Livnat, Adi, and Nicholas Pippenger. 2006. "An Optimal Brain can be Composed of Conflicting Agents." *Proceedings of the National Academy of Sciences*, 103(9): 3198-3202.

55. Loewenstein, George. 1996. "Out of Control: Visceral Influences on Behavior." *Organizational Behavior and Human Decision Processes*, 65(3): 272-292.

56. Loewenstein, George, and Edward O'Donoghue. 2005. "Animal Spirits: Affective and Deliberative Processes in Economic Behavior." Unpublished.

57. Loewenstein, George, and Drazen Prelec. 1991. "Negative Time Preferences." *American Economic Review papers and Proceedings*, 81(2): 347-352.

58. Loewenstein, George, Roberto Weber, Janine Flory, Stephen Manuck, and Matthew Muldoon. 2001. "Dimensions of Time-Discounting." Unpublished.

59. McClure, Samuel, David Laibson, George Loewenstein, and Jonathan Cohen. 2004. "Separate Neural Systems Value Immediate and Delayed Monetary Rewards." *Science*, 306(5695): 503-507.

60. Miller, Earl, and Jonathan Cohen. 2001. "An Integrative Theory of Prefrontal Cortex Function." *Annual Review of Neuroscience*, 24(3): 167-202.

61. Montague, Read, and Gregory Berns. 2002. "Neural Economics and the Biological Substrates of Valuation." *Neuron*, 36(2): 265-284.

62. Pascal, Blaise. 1670. *Les Pensées*. Penguin Classics: London, England.

63. Poldrack, Russell, and Paul Rodriguez. 2004. "How Do Memory Systems Interact? Evidence from Human Classification Learning." *Neurobiology of Learning and Memory*, 82(3): 324-332.

64. Rabin, Matthew. 1995. "Moral Preferences, Moral Constraints, and Self-Serving Biases." Unpublished.

65. Rauch, Scott, Paul Whalen, Cary Savage, Tim Curran, Adair Kendrick, Halle Brown, George Bush, Hans Breiter, and Bruce Rosen. 1997. "Striatal Recruitment during an Implicit Sequence Learning Task as Measured by fMRI." *Human Brain Mapping*, 5(2): 124-132.

66. Read, Daniel, George Loewenstein, and Matthew Rabin. 1999. "Choice Bracketing." *Journal of Risk and Uncertainty*, 19(1): 171-197.

67. Robinson, Terry, and Kent Berridge. 2003. "Addiction." *Annual Review of Psychology*, 54: 25-53.

68. Samuelson, Paul. 1937. "A Note on Measurement of Utility." *Review of Economic Studies*, 4(2): 155-161.

69. Shefrin, Hersh, and Richard Thaler. 1988. "The Behavioral Life-Cycle Hypothesis." *Economic Inquiry*, 26(4): 609-643.

70. Simonson, Itamar. 1990. "The Effect of Purchase Quantity and Timing on Variety Seeking Behavior." *Journal of Marketing Research*, 27(2): 150-162.

71. Strotz, Robert. 1956. "Myopia and Inconsistency in Dynamic Utility Maximisation." *Review of Economic Studies*, 23(3): 165-180.

72. Thaler, Richard. 1985. "Mental Accounting and Consumer Choice." *Marketing Science*, 4(3): 199-214.

73. Thaler, Richard. 1990. "Anomalies. Saving, Fungibility, and Mental Accounts." *Journal of Economic Perspectives*, 4(1): 193-205.

74. Thaler, Richard, and Hersh Shefrin. 1981. "An Economic Theory of Self-Control." *Journal of Political Economy*, 89(2): 392-406.

75. Tooby, John, and Leda Cosmides. 1992. "The Psychological Foundations of Culture." In *The Adapted Mind: Evolutionary Psychology and the Generation of Culture*, ed. Jerome Barkow, Leda Cosmides and John Tooby, 19-136. Oxford: Oxford University Press.

76. Whalen, Paul, Scott Rauch, Nancy Etcoff, Sean McInerney, Michael Lee, and Michael Jenike. 1998. "Masked Presentations of Emotional Facial Expressions Modulate Amygdala Activity without Explicit Knowledge." *The Journal of Neuroscience*, 18(1): 411-418.

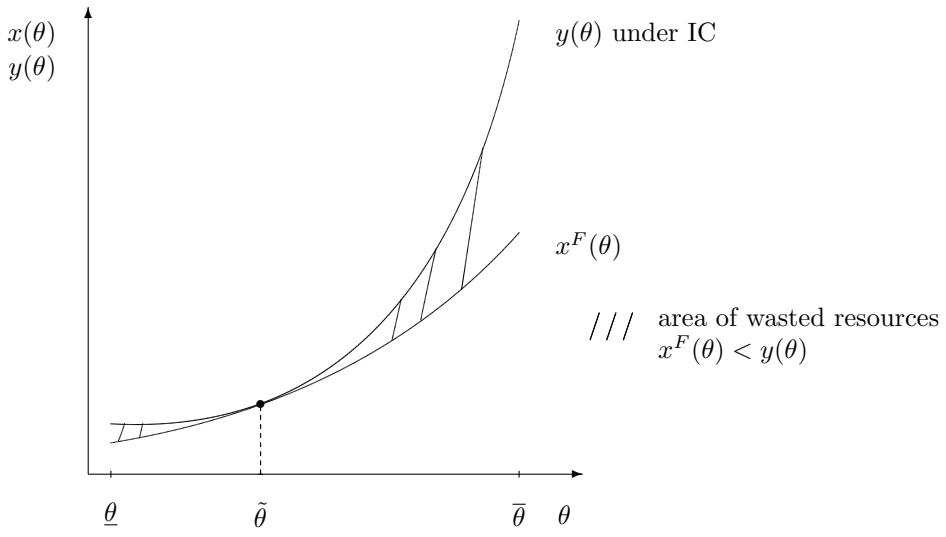

FIGURE 1A. OPTIMAL INCENTIVE COMPATIBLE CONTRACT

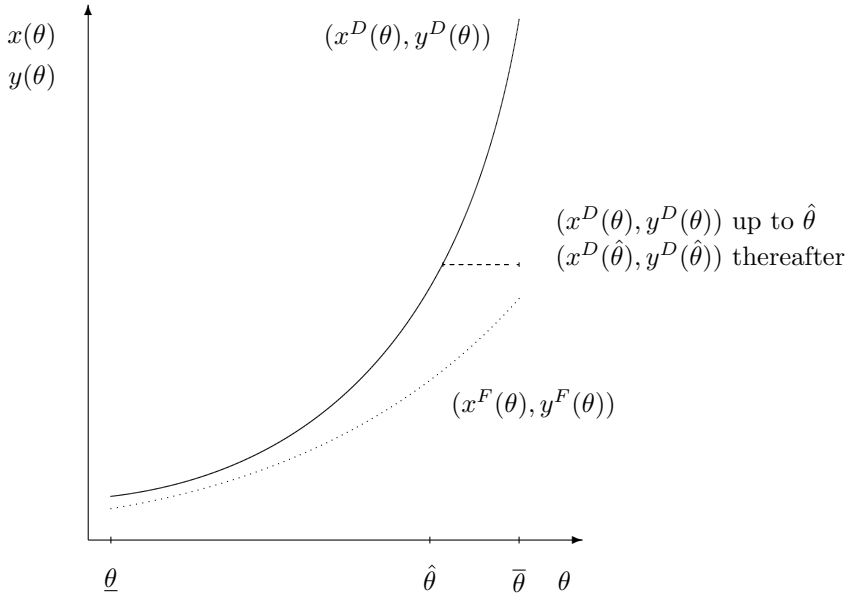

FIGURE 1B. FULL DELEGATION WITH AND WITHOUT CAP

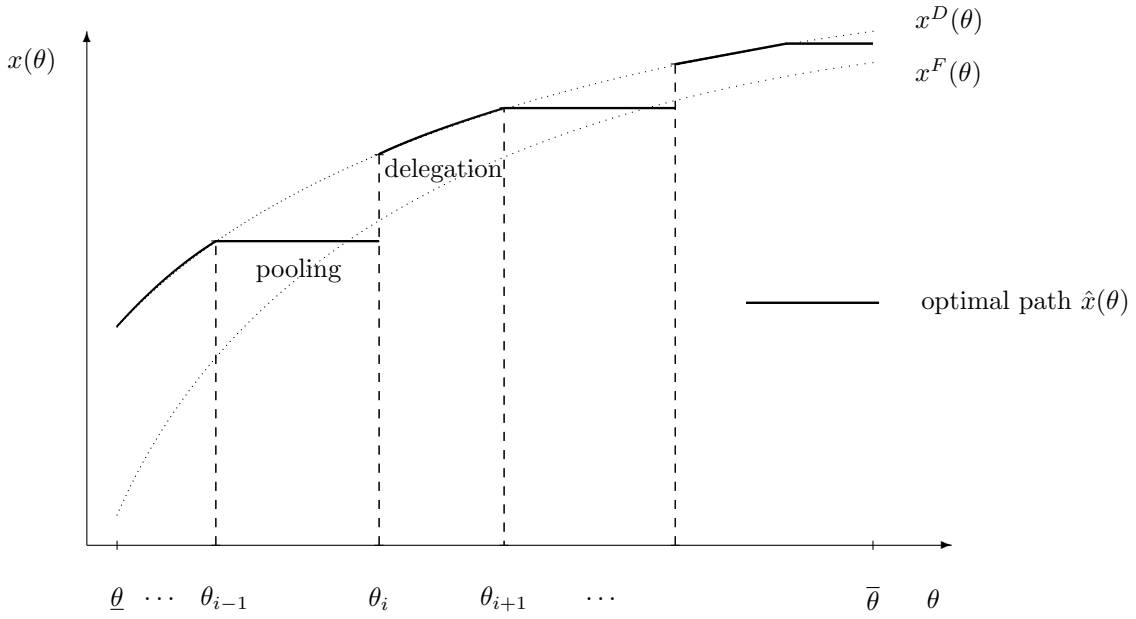

FIGURE 2. CONSUMPTION OF TEMPTING GOOD WHEN $T'' > 0$

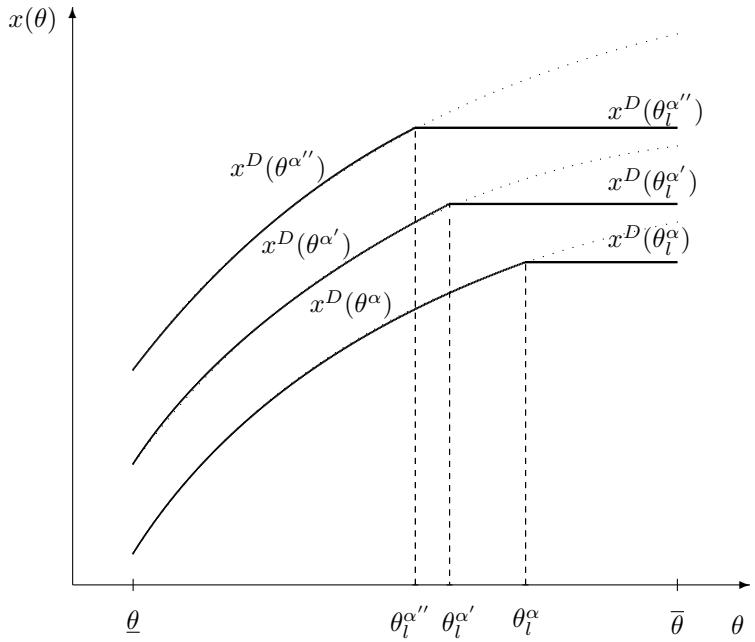

FIGURE 3. CONSUMPTION OF TEMPTING GOOD WITH $\alpha'' > \alpha' > \alpha > 1$