Games **2011**, *2*, 452-462; doi:10.3390/g2040452

OPEN ACCESS

games

ISSN 2073-4336

www.mdpi.com/journal/games

Article

The Resolution Game: A Dual Selves Perspective

Dimitri Migrow [1] and Matthias Uhl [2,*]

[1] Department of Economics, The University of Warwick, Coventry, CV4 7AL, UK;
E-Mail: d.migrow@warwick.ac.uk

[2] Max Planck Institute of Economics, Kahlaische Straße 10, 07745 Jena, Germany

* Author to whom correspondence should be addressed; E-Mail: uhl@econ.mpg.de;
Tel.: +49-3641-686-681; Fax: +49-3641-686-623.

Received: 18 July 2011; in revised form: 1 September 2011 / Accepted: 2 December 2011 / Published: 9 December 2011

Abstract: This article explains the emergence of an unique equilibrium resolution as the result of a compromise between two selves with different preferences. The stronger this difference is, the more generous the resolution gets. This result is in contrast to predictions of other models in which sinful consumption is distributed bimodally. Therefore, our result fits better with our daily observations concerning a lot of ambivalent goods where we often form nonrigid resolutions. The normative analysis uses the device of a hypothetical impartial self that regards both conflicting motives as equally legitimate. The result of this analysis is dilemmatic. It demonstrates that the resolution is broken too often to be welfare maximal. However, the introduction of external self-commitment devices results in their overuse and is welfare decreasing.

Keywords: dual selves; resolutions; self-control; impartiality

"You can make a resolution. This may be the most common way that people deal with temporary preferences but also the most mysterious." (Ainslie [1], p. 78)

1. Introduction

What would Ulysses do if there was no mast in the middle of the ship? He would be doomed as there is no other way to resist the enchanting songs of the Sirens. Most importantly, he cannot offer his curious self the possibility to listen to the songs "just a little". Highly addictive drugs, like heroin, are modern day equivalents to the songs of Sirens. However, the Ulysses metaphor may be overemphasized in the existing literature. Many goods or bads—depending on the self that makes the judgment—are truly ambivalent rather than instantly fatal or addictive. In their case, resolutions, a product of willpower, may quite frequently do the trick. It is the decision to consume these ambivalent goods that faces us every day. It is the consumption of these goods that our article focuses on. We all catch ourselves forming resolutions. They are: "I will not have more than three drinks tonight" rather than "I will try heroin just once". Resolutions have been largely disregarded in the existing literature (see also Ainslie [1]).

We argue, that along with the overemphasis of very extreme cases of intrapersonal conflict on the positive side of the analysis, there is a normative overemphasis of the planner's interest in the economic literature, which we discuss later in this text.

Our analysis uses the older idea of dual self models to explain a previously neglected but important phenomenon in human existence. An early dual self model was proposed by Thaler and Shefrin [2]. However, the question of how many selves a model needs to account for a subject's intrapersonal conflict was not resolved. Models of hyperbolic discounting, for example, constitute and endless string of selves over time. The dual self framework framework has recently gained support from research in both psychology and psychiatry (see, e.g., Frank [3]), from the literature on artificial intelligence and also from modern economic literature (see Ding [4] for the literature review). From a researcher's perspective, dual models are a simple and useful tool to model an intrapersonal conflict. On the normative side, the model regards each of the selves as equally legitimate agent within a person.

The most closely related literature to our topic is the literature on models of self-control. However, the focus there is once again on addiction. Bernheim and Rangel [5] analyze an addiction model with a hot mode and a cool self. Their cost structure is of zero-infinite type dependent on the mode and so differs from ours: our costs increase continuously with the extent of commitment and cant be infinite. Fudenberg and Levine [6] offer different mechanisms of self-control depending on setting. This might be the choice of the preferences for the myopic self by the long-run self, the limiting of the alternatives available to the short-run self or the long-run self incurring short-run costs to reduce future self control costs. As in Fudenberg and Levine, we also find a unique equilibrium, but explain the observed behaviour with a single mechanism and a simpler setting. Furthermore, we abstract from the misperception issues as discussed in O'Donoghue and Loewenstein [7].

The article proceeds as follows. Section 2 presents the basic model. Section 3 provides the results of our positive analysis. Section 4 presents our normative analysis. The article concludes with a discussion in Section 5.

2. Basic Model

Consider two players, a planning self, S_p, and an enjoying self, S_e. The identity of a person is formed by these two selves. Each of them takes control over the body at different points in time. S_p becomes active in a reflective state of planning, while S_e becomes active in a later state when consumption takes place. Each self determines the acting of the person in a given period of time.

2.1. Resolutions

The utility function of $S_i, i \in \{p, e\}$, is denoted by $u_i = u_i(\gamma)$, where γ is the consumption level[1].

S_p, as a Stackelberg leader, can improve S_e's well-being by making a generous resolution which allows S_e to consume a positive amount of the target good, $\gamma > 0$.

We assume that the utilities after making resolutions, $u_i(\gamma)$, are given by

$$u_i(\gamma) = (1 - \gamma)u_i + (\gamma - C(\gamma))\frac{u_p + u_e}{2} \tag{1}$$

We have chosen the above form because we think it helps to capture the idea of resolution as a form of "utility redistribution" between the two selves. Note, however, that the assumption of concavity for S_e is essential for comparative statics which follows in Section 3.

The first term of Equation (1) captures the perceived drawbacks of consumption. These could be moral scruples, concerns about negative health effects or concerns about increasing debts. The perceived pleasure of consumption is captured by the second term of Equation (1). Here, we may think of flavour, comfort or excitement.

We assume that $u_p(\gamma = 0) > 0$ and $u_e(\gamma = 0) = 0$: this captures the idea that S_p receives positive utility by non-consumption of an ambivalent good whereas S_e derives utility only from consumption of such good.

Note that with these basic specifications the relative weights that the two selves give to the drawbacks and pleasures of consumption in Equation (1) are asymmetric. Pleasure is weighted relatively higher by S_e, while the drawbacks are weighted relatively higher by S_p. Again, the intuition here is that S_e is the person in a more affective state where she simply enjoys consumption, while S_p is the person in a more reflective state.

As self-control requires sustainment of attention to a goal formulated by S_p, we introduce attention costs, $C(\gamma)$. Costs of attention are consistent with psychological literature which basically interprets self-control as a kind of attention control (see, e.g., Benhabib and Bisin [8]). We assume that $C(\gamma) \geq 0$, $\gamma \geq 0$, $\gamma \geq C(\gamma)$ and $\frac{dC}{d\gamma}$, $\frac{d^2C}{d\gamma^2} > 0$.

2.2. Breach of Resolution

The action variable of S_e concerns the breach of resolution and is denoted by $\rho \in \{0, 1\}$.

If S_e adheres to the resolution, $\rho = 0$, utilities, $u_i(\rho = 0)$, are given by Equation (1).

If S_e breaks the resolution, $\rho = 1$, utilities, $u_i(\rho = 1)$, are given by

$$u_e(\rho = 1) = (1 - \mu)(u_p + u_e) \tag{2}$$

and

$$u_p(\rho = 1) = 0 \tag{3}$$

μ is a random variable which defines S_e's inhibition threshold to break the resolution. Let $f(\mu)$ be the density function of μ, where $\int_{\mu=0}^{1} f(\mu)d\mu = 1$. S_p knows the distribution of the inhibition threshold, $f(\mu)$. Its actual realization can only be observed by S_e though.

[1] To save notations, in the following we denote $u_i(0)$ by u_i.

The very nature of a resolution is that it sets up an inhibition threshold which is psychologically costly to overcome. If restraint could simply be put aside, resolutions would not be credible and would therefore be meaningless. Making a resolution and setting up an inhibition threshold are two sides of the same coin. The intuition behind a stochastic inhibition threshold is that one cannot judge precisely how strong the temptation to break a resolution will be in a future situation. In this respect, one can sample from experience: one will usually have an idea about the distribution of the inhibition threshold. For example, one might know which people will be around at a party and how likely it is that they will try to convince one to break the resolution. In any case, the realization of the inhibition threshold is beyond S_p's control.

2.3. Timing

The sequence of events in the game is as follows:

(1) In $t = 0$, S_p decides about the generosity of its resolution, $\gamma \in [0, 1)$.
(2) In $t = 1$, an environmental shock is realized, $\mu \in [0, 1]$.
(3) In $t = 2$, S_e observes the realization of μ and decides whether to adhere to the resolution or whether to break it, $\rho \in \{0, 1\}$.
(4) In $t = 3$, both agents receive their payoffs.

3. Positive Analysis

To obtain S_p's optimal resolution, let us first consider the range of μ where S_e adheres to the resolution. This range follows from the generosity of the resolution, $\gamma \in [0, 1)$, made by S_p. A resolution is adhered to if $u_e(\rho = 0) \geq u_e(\rho = 1)$, implying that

$$(1-\gamma)u_e + \frac{(\gamma - C(\gamma))(u_p + u_e)}{2} \geq (1-\mu)(u_p + u_e)$$

or, using Equation (1),

$$\mu \geq (1-\gamma)\theta + \frac{\gamma + C(\gamma)}{2} \tag{4}$$

where $\theta \equiv \frac{u_p(\gamma=0)}{u_p(\gamma=0)+u_e(\gamma=O)}$.

Thus, we can define a critical value $\widetilde{\mu}$:

$$\widetilde{\mu} = (1-\gamma)\theta + \frac{\gamma + C(\gamma)}{2} \tag{5}$$

where S_e is indifferent between adhering to and breaking the resolution. The crucial point is that S_p can influence this critical value via the generosity of the resolution it chooses.

The first and second derivative of Equation (5) are given by

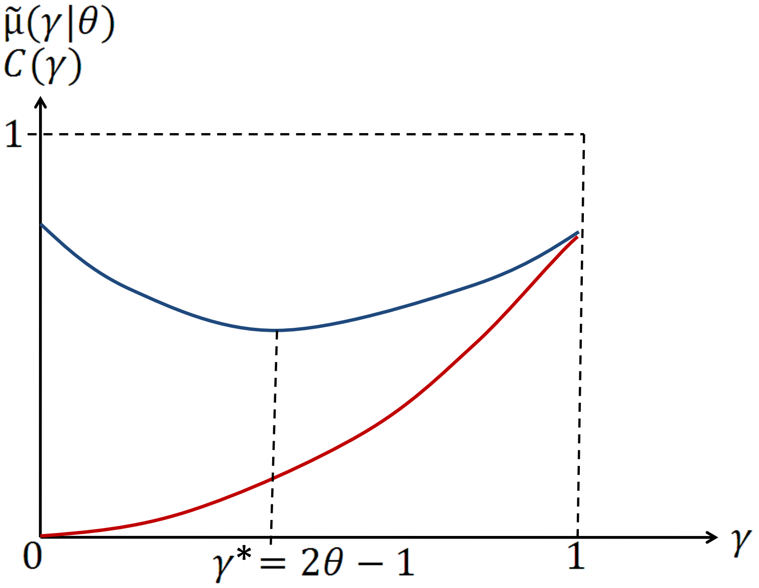
$$\frac{d\widetilde{\mu}}{d\gamma} = -\theta + \frac{1}{2} + \frac{1}{2}\frac{dC(\gamma)}{d\gamma} \tag{6}$$

and

$$\frac{d^2\widetilde{\mu}}{d\gamma^2} = \frac{1}{2}\frac{d^2C(\gamma)}{d\gamma^2} \tag{7}$$

From Equation (6) we can see that the critical threshold can be decreased by more generous resolutions until γ equals γ^*, which is implicitly given by $\frac{dC(\gamma^*)}{d\gamma} = 2\theta - 1$. From this level on, making more generous resolutions increases the critical value as the convexity of $C(\gamma)$ dominates. $\widetilde{\mu}$ is therefore U-shaped in γ. This is illustrated by the blue curve in Figure 1. The red curve represents the attention costs. If μ takes on a value below the blue curve, S_e will break the resolution.

Figure 1. Critical inhibition threshold.



S_p's utility is

$$u_p(\gamma) = \begin{cases} 0, & \text{if } \mu \in [0, \widetilde{\mu}) \\ (1-\gamma)u_p + \frac{\gamma - C(\gamma)}{2}(u_p + u_e), & \text{if } \mu \in [\widetilde{\mu}, 1] \end{cases} \tag{8}$$

Given Equations (5) and (8), we can derive S_p's optimal resolution.

Proposition 1. *The unique optimal resolution, $\widehat{\gamma}$, made by S_p in the resolution game is given by*

$$\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1 - F(\widetilde{\mu}(\widehat{\gamma}|\theta))} = \frac{\theta + \frac{\widehat{\gamma}-1}{2}}{\left(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2} - (1-\widehat{\gamma})\theta\right)\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma}} \tag{9}$$

S_p's optimal resolution is strictly smaller than the resolution minimizing the critical inhibition threshold, $\widehat{\gamma} < \gamma^*$. If the divergence between selves, θ, increases, S_p's optimal resolution increases as well.

Proof. See appendix.

4. Normative Analysis

4.1. The Impartial Hypothetical Self

Previously, we argued that the emergence of resolutions can be explained by the existence of two selves which represent conflicting motives. The question we will address in the following section is which "fair compromise" one would choose by an impartial observer. For this purpose, we introduce a third impartial hypothetical self which is not involved in the psychological conflict between S_p and S_e, but observes the conflict. It identifies both conflicting motives as manifest, in the sense that they take place systematically in certain situations. If we accept that preferences are revealed by actual choices or verbal and nonverbal expressions of satisfaction in each situation, we can ascribe two situational utilities to the person. The hypothetical self represents a preference, which, by definition, will only hold in moments where the person forces a special impartial attitude upon herself (see Harsanyi [9], p. 315)[2]. In these moments, she reasons behind a veil of ignorance and imagines that she has equal chance of being either in the planning self or the enjoying self. Each situation would be an "uncertain" prospect. Harsanyi ([9], p. 316) points out that if there is any objective criterion for comparing the utilities of individuals, the social welfare function will represent the sum of the individual utilities. He adds that intrapersonal utility comparisons are less problematic than interpersonal ones, since "the use of these indicators for comparing the utilities that a *given* person ascribes to different situations is relatively free of difficulty" (Harsanyi [9], p. 317). Adapting his argument for social welfare considerations, it seems natural for us to assume that the hypothetical self will give equal weights to both selves involved in conflict[3]. As will be outlined in the discussion section, this view contrasts with the hyperbolic discounting framework which which is biased towards the planning self.

4.2. A World Without External Self-Commitment

In the world of resolutions which we previously discussed, external self-commitment devices were not available. The equilibrium we obtain is therefore the result of self-regulation without the involvement of third parties. Let us now consider the welfare properties of this outcome. As argued before, the impartial hypothetical self defines the welfare of the person, U, as the sum of the utilities of both selves.

If S_e adheres to the resolution, $\rho = 0$, U is specified by the sum of $u_p(\gamma)$ and $u_e(\gamma)$ according to Equation (2),

$$U(\rho = 0) = (1 - C(\gamma))(u_p + u_e) \tag{10}$$

If S_e breaks the resolution, $\rho = 1$, U is specified by the sum of Equations (3) and (4), which is simply S_e's utility because S_p's utility is zero,

$$U(\rho = 1) = (1 - \mu)(u_p + u_e) \tag{11}$$

Thus, it is clear that for any given resolution, it is welfare maximizing that S_e adheres to it if $U(\rho \ = \ 0) \ \geq \ U(\rho = 1)$, implying

$$(1 - C(\gamma))(u_p + u_e) \geq (1 - \mu)(u_p + u_e)$$

or

$$\mu \geq C(\gamma) \tag{12}$$

[2] Note that the hypothetical self is different from the planning self because it is detached in a completely different sense. While the planning self is detached from the pleasures of consumption, the hypothetical self sees the pleasures of consumption as well as their remedies.

[3] Although we regard our weighting criteria as natural in terms of an impartial evaluation of welfare, there might be alternative weighting criteria: for example, using the amount of time in which the planning and the enjoying self "occupy" the mind of a person, as suggested by one of the referees.

If the inhibitions against breaking a resolution are weakly larger than $C(\gamma)$ it is beneficial to adhere to the resolution. Conversely, if the inhibitions are strictly weaker, it is beneficial to break the resolution. The red curve in Figure 1 represents the costs. If μ takes on a value below this curve, it is welfare enhancing to break the resolution. Since this is true for any resolution, it is true for S_p's optimal resolution, $\widehat{\gamma}$.

Proposition 2. *In equilibrium, a resolution is broken more frequently than welfare maximal.*

Proof. Actually, any given resolution will be adhered to by S_e if Equation (4) holds. Equation (4) can be rewritten as

$$\mu \geq \alpha C(\gamma) + (1-\gamma)\theta,\ \alpha \equiv \frac{1}{2} + \frac{1}{2}\frac{\gamma}{C(\gamma)} \geq 1 \tag{13}$$

Comparing the welfare criterion in Equation (12) with the actual criterion in Equation (13), it becomes evident that the resolution is broken more frequently than is welfare maximal since the right-hand side of Equation (13) is larger for any $\gamma < 1$. This is also true for $\widehat{\gamma}$ since $\widehat{\gamma} < 2\theta - 1 < 1$.

4.3. A World With External Self-Commitment

In a world without external self-commitment devices, S_p will choose a resolution, $\widehat{\gamma}$, that maximizes utility. Given the previous suboptimality result, the question of whether a welfare maximum could be achieved with external self-commitment devices arises. The possibility of external self-commitment, $\sigma \in \{0,1\}$, results in an enrichment of S_p's action space. The use of S_p's additional option restricts the action space of S_e to the singleton $\rho \in \{0\}$. S_p therefore eliminates the possibility of a breach of resolution by S_e when using this option. If a resolution will be enforced by external self-commitment, S_p will always form the most rigid resolution of abstinence, $\gamma = 0$. This is true because the whole point of a generous resolution is to make its breach less likely. Since a breach of resolution is excluded by an external self-commitment device by assumption, S_p will enforce its preferred consumption level of the target good, which is zero due to different utility levels by non-consumption of an ambivalent good and due to $\gamma \geq 0$ and $C(\gamma) \geq 0$.

S_p will choose to use a self-commitment device, $\sigma = 1$, if its utility from external self-commitment is larger than the expected utility from its optimal resolution, $\widehat{\gamma}$. If $\widehat{\gamma}$ offers an expected utility which is at least as high, it will abstain from using external self-commitment, $\sigma = 0$. S_p will therefore use external self-commitment, $\sigma = 1$, if $u_p(\sigma = 1) > u_p(\sigma = 0) = \mathbb{E}(u_p(\widehat{\gamma}))$, implying

$$(1-\phi)u_p > \left((1-\widehat{\gamma})u_p + \frac{(\widehat{\gamma} - C(\widehat{\gamma}))}{2}(u_p + u_e)\right)(1 - F(\widetilde{\mu}(\widehat{\gamma}|\theta)))$$

or

$$\widehat{\gamma} - \frac{\widehat{\gamma} - C(\widehat{\gamma})}{2\theta} + \left(\frac{\widehat{\gamma} - C(\widehat{\gamma})}{2\theta} + (1-\gamma)\right) F(\widetilde{\mu}(\widehat{\gamma}|\theta)) > \phi \tag{14}$$

where ϕ is the exogenously given cost of the external self-commitment device. Giving away car keys to prevent oneself from drinking is an example of external self-commitment (see Elster [10], p. 66). While resolutions come at the cost of attention efforts, external self-commitment devices have costs because third parties have to get involved. Examples include the psychological costs of embarrassment or a reputation loss due to admitting self-control problems. Monetary costs are less common but also possible: Bloom [11], for example, suggests to remove one's Internet cable and FedEx it to oneself to have a day of work without online distractions. FedExing incurs a monetary cost, of course. If the external self-commitment device is comparatively cheap, *i.e.*, Equation (14) holds, S_p will choose to use this device. Otherwise, S_p will form its optimal resolution, $\widehat{\gamma}$, which then gives it an expected utility that is at least as high.

Assume Equation (12) holds. Here, it will be welfare maximizing if S_e adheres to the resolution, $\rho = 0$. In this case, the external self-commitment alternative will be welfare dominating if $U(\sigma = 1) > U(\sigma = 0) = U(\rho = 0)$, implying

$$(1 - \phi)(u_p + u_e) > (1 - C(\gamma))(u_p + u_e)$$

or

$$C(\gamma) > \phi \tag{15}$$

Assume now Equation (12) does not hold. Here, it will be welfare maximizing if S_e breaks the resolution, $\rho = 1$. In this case, the external self-commitment alternative will be welfare dominating if $U(\sigma = 1) > U(\sigma = 0) = U(\rho = 1)$ implying

$$(1 - \phi)(u_p + u_e) > (1 - \mu)(u_p + u_e)$$

or

$$\mu > \phi \tag{16}$$

Proposition 3. *In equilibrium, the external self-commitment device is used more frequently than welfare maximal.*

Proof. Actually, S_p uses an external self-commitment device if Equation (14) holds. In the case when Equation (12) holds, by comparing Equation (14) and Equation (15), we can see that external self-commitment is used too frequently from a welfare maximizing perspective. This is true since the left-hand side of the welfare maximizing criterion in Equation (15) is smaller than the left-hand side of the actual criterion in Equation (14). To see this, note that $\frac{2\theta\gamma - \gamma + C(\gamma)}{2\theta} \geq C(\gamma)$, assuring that the right-hand side of Equation (14) is bigger than $C(\gamma)$. In the case where Equation (12) does not hold, the left-hand side of Equation (16) is even smaller since $C(\gamma) > \mu$ by assumption. Therefore it is clear that external self-commitment is used even more frequently than in the first case, which moves us further away from the welfare maximizing frequency of use. Since this is true for any resolution, it is also true for the optimal resolution, $\widehat{\gamma}$, that S_p will choose.

5. Discussion

Gul and Pesendorfer [12] criticize the multiple selves construct incorporated in quasi-hyperbolic discounting models for normative reasons. Since a self is identified with stable preferences and since preferences of hyperbolic discounters change over time, multiple selves are implied by their volatile decisions. Hyperbolic discounters constitute an endless string of time-indexed selves because choices depend on the time at which they are made. In this sense, multiple selves are an implication of hyperbolic discounting models rather than an explicit assumption. The additional discount parameter β in quasi-hyperbolic discounting models is therefore often interpreted as an "immediacy bias" and the resulting choice is seen as "irrational" (see Bernheim and Rangel [13]). Actually, this turns out to be a standard interpretation in the literature (see, e.g., O'Donoghue and Rabin [14], Gruber and Köszegi [15]). Usually, paternalistic arguments evolve that are motivated by the claim that β should be equal to one (see Gul and Pesendorfer [12], p. 31). Bernheim and Rangel [13] give a justification of the welfare criterion which sets β equal to one and thus basically equates the welfare of a person with her long-run preferences. If the consumer's time horizon becomes sufficiently long and she judges trade-offs between period t and $t + 1$ by exactly the same criteria in all but one period, the influence of any one "self" must vanish (Bernheim and Rangel [13], p. 14).

In other words, the selves in hyperbolic discounting models are not seen as competing and legitimate interests within a person but as a deviation from rationality. The exponential discounter becomes the normative benchmark. Cowen [16] criticizes a general bias towards long-run interests in the discussion on self-control problems in economics. Read ([17], pp. 685–686) raises the very valid point that not all intrapersonal conflict is due to hyperbolic discounting. This is the idea that the acting agent, or the enjoying self in our model, is better informed about the local circumstances of a certain choice than the planning agent. In models of hyperbolic discounting, it always holds that the planning self will make the "right" choice. This is the case, even though S_p does not know the future situation in which its plan is to be executed. In practice, it may be impossible to know when myopia is the cause of choice inconsistency or when only the acting agent realizes how appealing indulgence is (see Read [17], p. 686).

We argue that the dual selves perspective is more egalitarian from this perspective. Each self is explicitly identified with its own utility function that captures its own motives. Biases are mutual and the term "irrational" becomes vacuous. If two conflicting and recurrent main drives can be identified within a person in the context of resolutions both are ascribed their own "rationale". If two (or more) utility functions are formulated whose interaction describes consumer behavior, any aggregation rule is necessarily a welfare criterion. Thus, value judgments materialize and become vulnerable instead of sneaking in through the back door.

Acknowledgements

We thank Eberhard Feess, Werner Güth, David Hugh-Jones, Vittoria Levati, David Levine, Christoph Vanberg and Tony Xiao Lin for comments. We are especially thankful to Birendra Rai for extensive comments.

References

1. Ainslie, G. *Breakdown of Will*; Cambridge University Press: Cambridge, UK, 2001.
2. Thaler, R.; Shefrin, H. An economic theory of self-control. *J. Polit. Econ.* **1981**, *89*, 392–406.
3. Frank, B. The use of internal games: The case of addiction. *J. Econ. Psychol.* **1996**, *17*, 651–660.
4. Ding, M. A theory of intraperson games. *J. Mark.* **2007**, *71*, 1–11.
5. Bernheim, B.; Rangel, A. Addiction and cue—Triggered addiction processes. *Am. Econ. Rev.* **2004**, *94*, 1558–1590.
6. Fudenberg, D.; Levine, D. A dual-self model of impulse control. *Am. Econ. Rev.* **2006**, *96*, 1449–1476.

7. Loewenstein, G.; O'Donoghue, T. *Animal Spirits: Affective and Deliberative Processes in Economic Behavior*; SSRN Working Paper; Ithaca, NY, USA, 2004. Available online: http://ssrn.com/abstract=539843 (accessed on 2 December 2011).
8. Benhabib, J.; Bisin, A. Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games Econ. Behav.* **2005**, *52*, 460–492.
9. Harsanyi, J. Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *J. Polit. Econ.* **1955**, *63*, 309–324.
10. Elster, J. *Ulysses Unbound*; Cambridge University Press: Cambridge, UK, 2000.
11. Bloom, P. *First Person Plural*; The Atlantic: Washington, DC, USA, November 2008. Available online: http://www.theatlantic.com/magazine/archive/˜2008/11/first-person-plural/7055/ (accessed on 1 October 2011).
12. Gul, F.; Pesendorfer, W. The case for mindless economics. In *The Foundations of Positive and Normative Economics. A Handbook*; Caplin, A., Schotter, A., Eds.; Oxford University Press: Oxford, UK, 2008; pp. 3–39.
13. Bernheim, D.; Rangel, A. Behavioral public economics: Welfare and policy analysis with nonstandard decision-makers. In *Behavioral Economics and Its Applications*; Diamond, P., Vartiainen, H., Eds.; Princeton University Press: Princeton, NJ, USA, 2007; pp. 7–77.
14. O'Donoghue, T.; Rabin, M. Studying optimal paternalism, illustrated by a model of sin taxes. *Am. Econ. Rev.* **2003**, *93*, 186–191.
15. Gruber, J.; Köszegi, B. Tax incidence when individuals are time-inconsistent: The case of cigarette excise taxes. *J. Public Econ.* **2004**, *88*, 1959–1987.
16. Cowen, T. Self-constraint versus self-liberation. *Ethics* **1991**, *101*, 360–373.
17. Read, D. Which side are you on? The ethics of self-command. *J. Econ. Psychol.* **2006**, *27*, 681–693.

Appendix

The expected utility of S_p is

$$\begin{aligned}\mathbb{E}(u_p(\gamma|\theta)) &= \left((1-\gamma)u_p + \frac{(\gamma - C(\gamma))}{2}(u_p+u_e)\right)\int_{\widetilde{\mu}(\gamma|\theta)}^{1} f(\mu)d\mu \\ &= \left((1-\gamma)u_p + \frac{(\gamma - C(\gamma))}{2}(u_p+u_e))(1-F(\widetilde{\mu}(\gamma|\theta))\right)\end{aligned}$$

Optimization calculus over γ implies

$$\frac{d}{d\gamma}\mathbb{E}(u_p(\gamma|\theta)) = 0$$

which means

$$\begin{aligned}\frac{d}{d\gamma}\,(1-\gamma)u_p(1-F(\widetilde{\mu}(\gamma|\theta))) + \frac{d}{d\gamma}\,\frac{\gamma}{2}(u_p+u_e)(1-F(\widetilde{\mu}(\gamma|\theta))) \\ -\frac{d}{d\gamma}\,\frac{C(\gamma)}{2}(u_p+u_e)(1-F(\widetilde{\mu}(\gamma|\theta))) = 0\end{aligned}$$

or

$$
\begin{aligned}
-u_p(1-F(\widetilde{\mu}(\gamma|\theta)))-(1-\gamma)u_pf(\widetilde{\mu}(\gamma|\theta))\frac{\partial\widetilde{\mu}(\gamma|\theta)}{\partial\gamma}\\
+\frac{1}{2}(u_p+u_e)(1-F(\widetilde{\mu}(\gamma|\theta)))-\frac{\gamma}{2}(u_p+u_e)f(\widetilde{\mu}(\gamma|\theta))\frac{\partial\widetilde{\mu}(\gamma|\theta)}{\partial\gamma}\\
-\frac{\gamma}{2}(u_p+u_e)(1-F(\widetilde{\mu}(\gamma|\theta)))+\frac{C(\gamma)}{2}(u_p+u_e)f(\widetilde{\mu}(\gamma|\theta))\frac{\partial\widetilde{\mu}(\gamma|\theta)}{\partial\gamma}=0
\end{aligned}
$$

Dividing the above equation by $(u_p + u_e)$ and rearranging it yields:

$$
\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1-F(\widetilde{\mu}(\widehat{\gamma}|\theta))}=\frac{\theta+\frac{\widehat{\gamma}-1}{2}}{\left(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2}-(1-\widehat{\gamma})\theta\right)\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\gamma}}
$$

Given the positive left-hand side of the above expression, note that the right-hand side is positive if $\widehat{\gamma} < \gamma^*$. While $\left(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2}-(1-\widehat{\gamma})\theta\right)$ is always negative since $\gamma \geq C(\gamma)\ \forall\gamma$ by assumption, the right-hand side is positive only in the case where $\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma} < 0$. Given the U-shaped form of $\widetilde{\mu}(\gamma|\theta)$ with $\widetilde{\mu}(\gamma^*|\theta) < \widetilde{\mu}(\gamma'|\theta)\ \forall\gamma' \neq \gamma^*$, clearly $\frac{d\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\gamma} < 0$ only if $\gamma < \gamma^*$. Thus, $\widehat{\gamma} < \gamma^*$.

Since $\frac{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))}{\partial\widehat{\gamma}} \neq 0,\ F(\widetilde{\mu}(\widehat{\gamma}|\theta))$ defines $\widehat{\gamma}$ as an implicit function of θ. The question is now how $\widehat{\gamma}$ changes if θ changes marginally. By total differentiation we get

$$
\frac{d\widehat{\gamma}}{d\theta}=-\frac{\frac{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))}{\partial\theta}}{\frac{\partial F(\widetilde{\mu}(\widehat{\gamma}|\theta))}{\partial\widehat{\gamma}}}
$$

or

$$
\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\widehat{\gamma}}=-\frac{\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{\partial\theta}}{\frac{d\widehat{\gamma}}{d\theta}}
$$

Substitution yields:

$$
\frac{f(\widetilde{\mu}(\widehat{\gamma}|\theta))}{1-F(\widetilde{\mu}(\widehat{\gamma}|\theta))}=\frac{(\theta+\frac{\widehat{\gamma}-1}{2})\frac{d\widehat{\gamma}}{d\theta}}{-\left(\frac{C(\widehat{\gamma})-\widehat{\gamma}}{2}-(1-\widehat{\gamma})\theta\right)\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta}{\partial\theta}}
$$

Since $\frac{\partial\widetilde{\mu}(\widehat{\gamma}|\theta)}{d\theta} > 0$, the above equation shows that $\frac{d\widehat{\gamma}}{d\theta} > 0$.