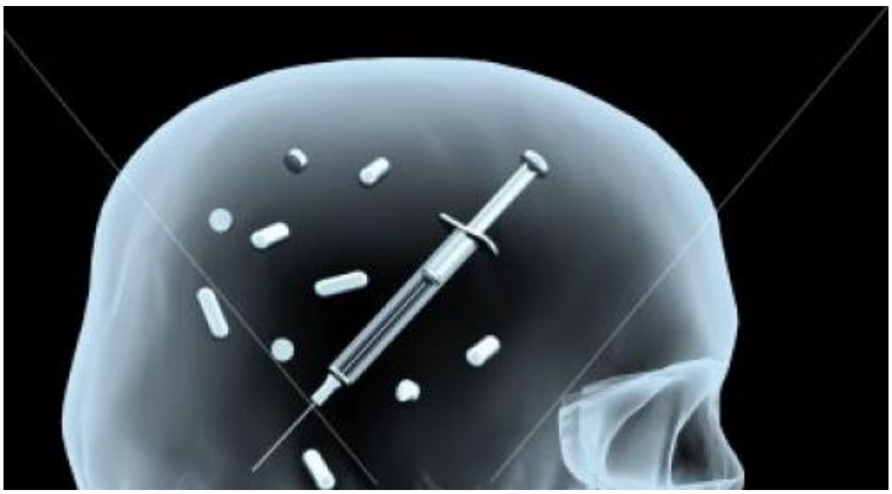
# A Theory of Rational Addiction

Gary S. Becker and Kevin M. Murphy (*JPE*, 1988)



Notes by Team Grossman (Spring 2015)
Meagan Madden and Mavzuna Turaeva

## Contents



## Introduction

The model of addiction by Becker and Murphy (B&M) (1998) was one of the first to model the behavior of addicts in a "rational" way. By rational, they mean that addicts have stable preferences and make utility-maximizing decisions about whether or not to consume an addictive good, and they are capable of taking into account the consequences of the current consumption on the future. This approach differed from previous models of addiction which assumed that addicts are "myopic" and heavily discount (or ignore) utility/disutility in future periods.

B&M build on a few earlier economic models of tastes and addiction, including Stigler and Becker (1997) and Iannaccone (1984, 1986). In "Rational Addiction," B&M formalize the definition of an addictive good as one where your consumption in one period is complementary to consumption in the next period (i.e. adjacent complementarity). This definition encompasses addictions that are harmful or beneficial.

Therefore, we can think of many types of addictive goods in this model, including cigarettes, alcohol, work, social media, food, and exercise.

In the B&M model, individuals choose between spending money on addictive goods and spending money on all other goods. Addictive goods differ from other goods because utility from current consumption depends on consumption in other periods, and these inter-temporal effects are contained in a "stock of consumption" variable. To capture the change in this stock over time and how it affects your consumption decision today, B&M develop a continuous time model. Instead choosing the optimal *consumption bundle* to maximize utility subject to the budget (static case), individuals choose the optimal *consumption paths* to maximize lifetime utility subject to the lifetime budget (dynamic case). An additional constraint in the dynamic case is the flow of the stock variable.

From this dynamic model, B&M predict consumption patterns of addicts, including price responsiveness in the short-run and long-run as well as binges and purges. There is strong empirical evidence to support several aspects of the rational addiction model. However, critics still debate whether it is appropriate to consider addicts in a rational thinking model.

## Key Notation

| Component | Description |
|---|---|
| $c(t)$ | Consumption of an addictive good which is related to $S(t)$ |
| $y(t)$ | Consumption of a non-addictive good |
| $S(t)$ | "consumption stock" which is related to your consumption of addictive goods in previous periods $t$ through "learning" |
| $D(t)$ | Expenditure on the endogenous depreciation of $S$ to "forget" your "consumption stock" of addictive goods (e.g. rehabilitation costs) |
| $\dot{S}(t)$ | Rate of change of $S(t)$ over time |
| $A_0$ | Initial assets |
| $e^{-\sigma t}$ | Discount factor with constant rate of time preference $\sigma$ |
| $t$ | Time |
| $u(\cdot)$ | Utility function a function of $y$, $c$, and $S$ (strongly concave in $y, c$ and $S$). |
| $r$ | Rate of interest, constant over time |
| $w(S)$ | Earnings as a concave function of the stock of consumption capital |
| $p(t)$ | Prices |
| $a(t)$ | The shadow price of $S$, i.e. the marginal utility of an additional unit of consumption stock $S$; also interpreted as the discounted utility and monetary cost/benefit of additional consumption of $c$ through the effect on $S$. |
| $\Pi_c(t)$ | The shadow price of $c(t)$, i.e. the marginal utility of an additional of $c(t)$; also interpreted as the full price of $c(t)$ which equals the sum of the market price of $c(t)$ plus the monetary value of the cost/benefit of future consumption |
| $V(\cdot)$ | Lifetime indirect utility (i.e. the maximum obtainable utility) |
| $\mu = \frac{\partial V}{\partial A_0}$ | Shadow price of wealth; i.e. the marginal utility of an additional unit of wealth |
| $\delta$ | Rate of depreciation of the consumption stock which is constant |
| $k$ | "Composite constant" of indirect utility that depends on $A_0, \mu, \sigma$ (enters the simplified dynamic optimization problem) |
| $S^*$ | Steady state value of $S$ |

## The Model

Utility is derived from consumption of goods $y$ and addictive goods $c$ as well as the "stock of addictive capital" $S$:

$$u(t) = u[y(t), c(t), S(t)] \tag{1}$$

Utility is separable over time in $y$, $c$ and $S$, but not in $y$ and $c$ alone. This occurs because your "stock" $S$ depends on your past consumption of addictive goods $c$. Equation (1) implies that the amount that utility you derive from consumption today partly depends on your "stock" which depends what you consumed in previous periods. Investment in this "stock" occurs through the following function:

$$\dot{S}(t) = c(t) - \delta S(t) - h[D(t)] \tag{2}$$

Total lifetime utility is the discounted $u(t)$ evaluated over continuous time, with constant rate of time preference $\sigma$. (See A1 for more on time preferences)

$$U(\cdot) = \int_0^T e^{-\sigma t} u[y(t), c(t), S(t)]dt \tag{3}$$

The budget constraint says that consumption of all choice variables at prices $\boldsymbol{p}(t)$ must be ≤ initial assets ($A_0$) and future earnings to capital stock, $w(S(t))$. The budget is defined in present value terms using interest rate $r$:

$$\int_0^T e^{-rt}\left[y(t) + p_c(t)c(t) + p_d(t)D(t)\right]dt \leq A_0 + \int_0^T e^{-rt} w(S(t))dt \tag{4}$$

Individuals want to choose the best consumption path in order to maximize their lifetime utility (3) subject to their lifetime budget (4). This is basically a control problem with control variables $c(t)$, $y(t)$, and $D(t)$; state variable $S(t)$, and the transition state function $\dot{S}(t)$. Using calculus of variations we setup the Hamilton and derive the FOCs for an extrema (See A2 for formal derivations):

$$u_y = \mu e^{(\sigma - r)t}$$

$$h_d(t) \cdot a(t) = \mu p_d(t) e^{-rt} \tag{5}$$

$$u_c(t) = \mu p_c(t) e^{-(\sigma - r)t} - e^{\sigma t} a(t)$$

Where $a(t)$ and $\mu$ are the shadow price of consumption stock and wealth, respectively. These *shadow prices* are interpreted at the marginal utility from small change in either stock or wealth.

## Dynamics

The solution to the dynamic optimization problem is a differential equation which determines the optimal consumption path of $c$ and $y$ and state transition path of $S$. To more easily find the solution, B&M simplify the model and then find the analytical and graphical solution to the differential equation.

Simplifying Assumptions:

a. Infinite time: $T = \infty$

b. Time preference equals interest rate: $\sigma = r$

c. No endogenous depreciation of $S$: $D(t) = 0$

d. Quadratic utility and earnings functions, so the relationship between $c$ and $S$ is linear.

Since B&M are most interested in the dynamics of $c$ and $S$, they optimize $y$ out of the value function. Essentially, this allows them to examine small deviations of $c$ and $S$ near the steady-states and easily represent the dynamics of these variables of interest using a two-dimensional phase diagram. Therefore, the value function is expressed as a quadratic of only $c$ and $S$ through a Taylor Series Expansion:

$$F(t) = \alpha_c c(t) + \alpha_s S(t) + \frac{\alpha_{cc}}{2}[c(t)]^2 + \frac{\alpha_{ss}}{2}[S(t)]^2 + \alpha_{cs} c(t)S(t) - \mu p_c c(t) \qquad (6)$$

Where the coefficients $\alpha_s$ and $\alpha_{ss}$ contain the marginal and secondary effects of $S$ on the utility function ($u$) and the earnings function ($w$). The table below depicts the first- and second-order conditions for the case of harmful goods and the case of beneficial goods.

<table>
<tr><th>Type</th><th>FOCs</th><th>Notation in (6)</th><th>Description</th><th>Second Order</th><th>Notation in (6)</th></tr>
<tr><td rowspan="2">Harmful addictive goods</td><td>$u_s < 0$</td><td rowspan="2">$\frac{\partial F}{\partial S} = \alpha_s < 0$</td><td>Adverse effect on utility from an increase in “consumption stock” of a harmful addictive good</td><td>$u_{ss} < 0$</td><td rowspan="2">$\frac{\partial^2 F}{\partial S^2} = \alpha_{ss} < 0$</td></tr>
<tr><td>$w_s < 0$</td><td>Adverse effect on earnings from an increase in “consumption stock” of a harmful addictive good</td><td>$w_{ss} < 0$</td></tr>
<tr><td rowspan="2">Beneficial addictive goods</td><td>$u_s > 0$</td><td rowspan="2">$\frac{\partial F}{\partial S} = \alpha_s > 0$</td><td>Positive effect on utility from “consumption stock” of beneficial addictive good</td><td>$u_{ss} < 0$</td><td rowspan="2">$\frac{\partial^2 F}{\partial S^2} = \alpha_{ss} < 0$</td></tr>
<tr><td>$w_s > 0$</td><td>Positive effect on earnings from “consumption stock” of beneficial addictive good</td><td>$w_{ss} < 0$</td></tr>
<tr><td rowspan="2">For all “addictive” goods</td><td rowspan="2">$u_c > 0$</td><td rowspan="2">$\frac{\partial F}{\partial c} = \alpha_c > 0$</td><td rowspan="2">⇒Utility from consumption of addictive goods is increasing, but at a decreasing rate<br>⇒ $u_{cs} > 0$ is the key element of the model which indicates complementarity of the consumption and capital stock; this means that the marginal utility of current consumption of an addictive good increases with past consumption of $c$[‡]</td><td>$u_{cc} < 0$</td><td>$\frac{\partial^2 F}{\partial c^2} = \alpha_{cc} < 0$</td></tr>
<tr><td>$u_{cs} > 0$</td><td>$\frac{\partial^2 F}{\partial c \partial S} = \alpha_{cs} > 0$</td></tr>
</table>

[‡]Alternatively, $u_{cs} < 0$ means $c$ and $S$ are substitutes, and $u_{cs} = 0$ means $c$ and $S$ are not related. Under these conditions $c$ is non-addictive.

The value function (6) gives the instantaneous indirect utility of your consumption choices ($c$, $S$) in time $t$. Individuals want to maximize instantaneous utility over all time, so the indirect lifetime utility from this maximization problem is given by

$$V(A_0, S_0, p_c) = k + \max_{c,S} \int_0^\infty e^{-\sigma t} F[S(t), c(t)]dt \qquad (7)$$

In (7), $k$ is a constant utility value that is a function of $A_0$, $\mu_2$, the $\sigma$ discount factor on $y$, plus the coefficients for $y$ in the quadratic utility function (i.e. the coefficients before B&M optimized out $y$).

Therefore, in the simplified dynamic model, your indirect lifetime utility comes from the optimal paths of $S$ and $c$ (which are maximized) and the optimal path of $y$ (which is already given in $k$). Individuals maximize (7) subject to the investment function for "consumption stock" ($\dot{S}$) given by (2) as well as a transversality condition:

$$\lim_{t\to\infty} e^{-\sigma t}\,[S(t)]^2 = 0 \tag{8}$$

This condition implies that your present value of a "consumption stock" must approach zero on an infinite horizon. We also know that $[c(t)]^2$ is bounded (See A4). This implies that the linear terms in (6) are similarly bounded.

We can solve the optimization problem above using calculus of variations in order to find the Euler Equation which is a differential equation determining the optimal paths of $c$ and $S$ (See A3 for derivation):

$$\ddot{S} - \sigma\dot{S} - BS = \frac{(\sigma+\delta)a_c + a_s}{a_{cc}} - \frac{(\sigma+\delta)p_c\mu}{a_{cc}} \tag{9}$$

Where $B = \delta(\sigma+\delta) + \frac{a_{ss}}{a_{cc}} + (\sigma+2\delta)\frac{a_{cs}}{a_{cc}}$. The ODE (9) has two roots of the form $\lambda = \frac{\sigma\pm\sqrt{\sigma^2+4B}}{2}$. The larger root violates the transversality condition (8) (See A4). Using the first root and the initial stock, we can find the optimal path of $S$ and also the optimal path of $c$:

$$S(t) = de^{\lambda_1 t} + S^* \text{, where } d = S_0 - S^* \tag{A4.2}$$

$$c(t) = \dot{S}(t) + \delta S(t) = (\delta+\lambda_1)S(t) - \lambda_1 S^* \tag{A4.3}$$

The relation between $c$ and $S$ is determined by the term $(\delta+\lambda_1)$ in (A4.3) which is determined by:

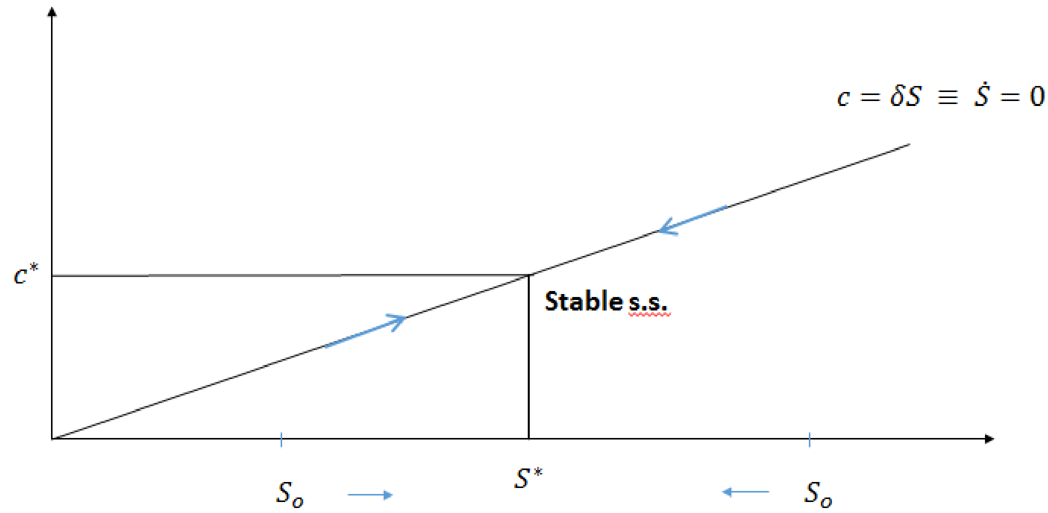
$$(\sigma+2\delta)\alpha_{cs} \gtreqless -\alpha_{ss} > 0 \tag{A5.1}$$

If the LHS is greater than the RHS in (A5.1), $\lambda_1 > -\delta$ and there is a positive relationship between past and present consumption. The other case $\lambda_1 < -\delta$ implies a negative relationship, and $\lambda_1 = -\delta$ implies that past and present consumption are unrelated.

In the steady state, the rate of growth of capital stock is zero ($\dot{S} = 0$). This occurs only if the consumption in the last period is equal to the depreciation of the stock ($c(t) = \delta S(t)$). We can draw phase diagrams to represent the transitional dynamics between $c$ and $S$ where steady states occur on the line $c(t) = \delta S(t)$.

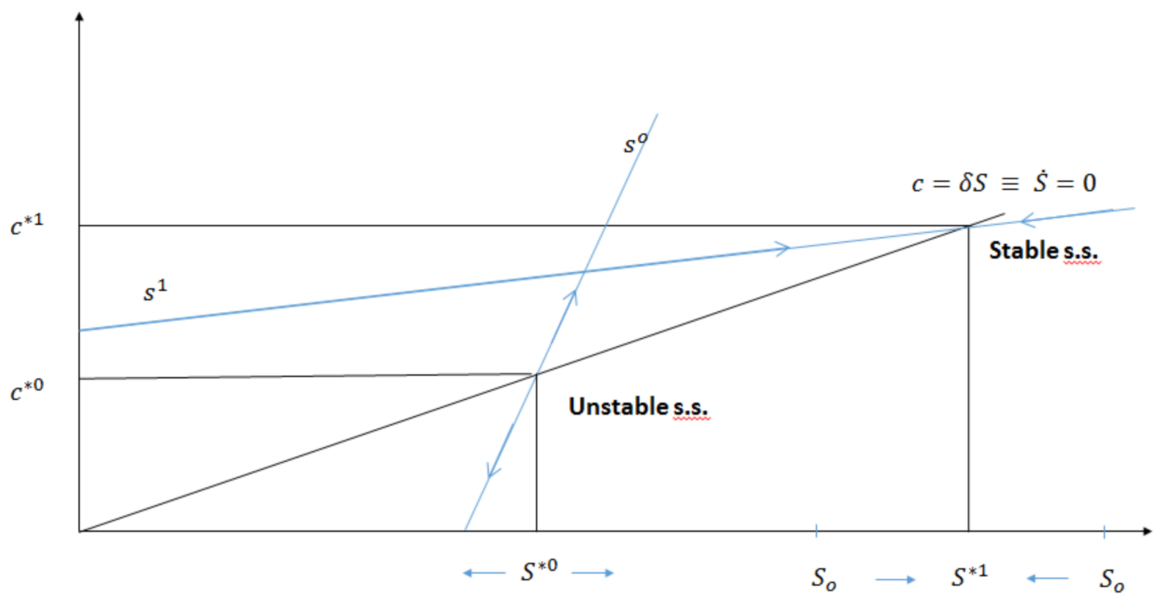
Fig. 1' below represents the simple case when the relationship between $S$ and $c$ is defined by a quadratic, and there is only 1 real root that solves the ODE and also satisfies transversality. If an individual is endowed with an initial level of stock ($S_0$), they will choose a level of $c$ on optimal path. Whether the individual starts to the left or right of $S^*$ does not matter; in either case the path will converge to the stable steady-state.

**Figure 1'.** Graphical solution of adjacent complementarity (simple case)




With a different form for $F(t)$, the solution will be different and other phase diagrams are possible, including multiple steady states, unsteady states, and jumps (See Fig. 1)

**Figure 1.** Graphical solution when cubic form for F(t) leads to two steady states, one stable and one not.




In Fig. 1, there are two steady states along the optimal path of the capital stock; one is stable and one is unstable. The optimal path line $s^1$ has stable steady state at $\delta S^{*1} = c^{*1}$. But the optimal path line $s^0$ has an unstable steady state at $\delta S^{*0} = c^{*0}$. The line $s^0$ cuts cuts the steady-state line from below.

An increase in parameters $\sigma$, $\delta$, or $a_{cs}$, will increase the degree of adjacent complementarity, which in turn makes the smaller root $\lambda_1$ large in algebraic value. The roots $\lambda_1$ and $\lambda_2$ will be positive when adjacent complementarity is strong enough to make $B < 0$ (See A5.) This makes the steady state unstable: consumption grows over time if initial consumption exceeds the steady state level, and it falls to zero when initial consumption is below the steady state level.

The unstable state is crucial in understanding of rational addictive behavior. An increase in the degree of potential addiction (i.e. the degree of complementarity $a_{cs}$) raises the likelihood that the steady state is unstable. Thus, unstable steady states can explain rational "pathological" addictions, in which the consumption of a good increases over time, even though the person fully anticipates future and his/her rate of time preference is no smaller than the rate of interest. The unstable steady state also helps in explain "normal" addictions that may involve rapid increases in consumption only for a while.

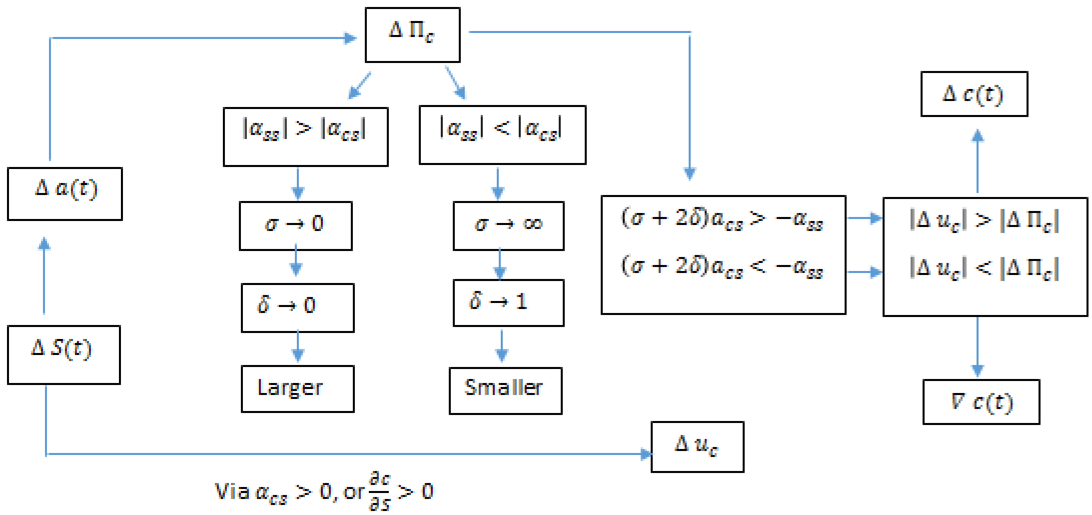
With two steady states relatively few people consistently consume small quantities of addictive goods. Consumption diverges from the unstable steady state either to zero or to a high level. Empirically, the distribution of consumption of highly addictive goods tends to be bimodal: people either consume a lot of it or very little (e.g. cigarettes or cocaine). On the other hand, alcohol consumption is not as addictive as nicotine or cocaine (for most people), so the distribution of alcohol consumption is more continuous.

## Adjacent Complementarity

Adjacent complementarity is the determining condition for addiction under rational addiction theory. For a good to be addictive, the consumption in adjacent periods must be complementarity. Mathematically, $(\sigma + 2\delta)\alpha_{cs} > -\alpha_{ss} > 0$ (See A5). The parameter $\alpha_{cs} > 0$ is interpreted as the degree of complementarity.

The concept of adjacent complementarity is closely related to the idea of reinforcement: that greater current consumption raises future consumption. Tolerance means that given levels of consumption are less satisfying when past consumption has been greater. Harmful addictions imply a form of tolerance because higher past consumption of harmful goods lowers present utility from the same level of consumption. B&M argue that the effect of tolerance exists only for harmful goods. For example, there is biological evidence of developing tolerance to drugs; after using drugs for a while, you do not receive the same "high" from the same amount of consumption. However, we think you can also develop a tolerance to beneficial addiction (e.g. runner's "high").

We see that different individuals have different addictions, and an inidividual may be addictive to some goods but not other. The parameters in the model ($\sigma$,$\delta$, plus all the $\alpha$ coefficients from (6)), determine the "addictiveness" of a good, and these parameters are different by person. For example, present-oriented individuals are potentially more addicted to harmful goods than future oriented individuals. The reason is that an increase in past consumption leads to a smaller rise in full price when the future is more heavility dicsounted. The folloing diagram captures the dynamics of consumption of the addictive good related to individually-determined parameters:




# Understanding Behaviors Related to Addictive Consumption

## Permanent Price Changes

B&M derive the own-price effect on demand for addictive goods, then take the partial derivative with respect to time:

$$\frac{\partial}{\partial t}\left[\frac{\partial c(t)}{\partial p_c}\right] = \frac{\partial \dot{c}}{\partial p_c} = \frac{\partial}{\partial p_c}\left(\frac{dc}{dS}\dot{S}\right) = \frac{dc}{dS}\frac{\partial \dot{S}}{\partial p_c} + \dot{S}\frac{\partial}{\partial p_c}\left(\frac{dc}{dS}\right) \tag{10}$$

In the steady-state $\dot{S} = 0$, so the term $\dot{S}\frac{\partial}{\partial p_c}\left(\frac{dc}{dS}\right) \approx 0$ near the steady-state. Therefore,

$$\frac{\partial}{\partial t}\left[\frac{\partial c(t)}{\partial p_c}\right] \approx \frac{dc}{dS}\frac{\partial \dot{S}}{\partial p_c} \tag{11}$$

The term $\frac{\partial \dot{S}}{\partial p_c}$ is necessarily negative because $\frac{\partial c}{\partial p_c} < 0$ (assuming normal goods), and $\dot{S}$ depends on $c$. Therefore, the sign of $\frac{\partial}{\partial t}\left[\frac{\partial c(t)}{\partial p_c}\right]$ is determined by $\frac{dc}{dS}$. If past and present consumption are complements, $\frac{dc}{dS} > 0$ and the effect of a price change increases over time.

Furthermore, B&M show that the long-term effect of price is often *greater* for *more addictive* goods. In the first case, assuming the quadratic form of $F(t)$ (6), the derivative of the FOCs yields:

$$\frac{dc^*}{dp_c} = \frac{\mu}{a_{cc}}\frac{\delta(\sigma+\delta)}{B} < 0 \tag{12}$$

The degree of addiction ($\alpha_{cS}$) lowers B in the denominator of (12), so that the change in steady state consumption w.r.t to price is greater for highly addictive goods.

Assuming another form on (6) could result in unsteady states which would further amplify the effect of price changes on consumption. B&M depict this in Fig. 2 with the curves $p^1p^1$ and $p^2p^2$ which have unstable steady states at $S^{1*}$ and $S^{2*}$, respectively. If your initial stock under $p^1$ is to the left of $S^{1*}$, then you will move toward zero consumption. If your initial stock under $p^2$ is to the right of $S^{2*}$, then you will move to the steady state of consumption where there is high stock and high consumption. Therefore, if your initial stock is between $S^{1*}$ and $S^{2*}$ and prices drop from $p^1$ to $p^2$, then your steady state will change. The effect of the price change moves your steady state value of consumption from zero to a much higher level of $c$.

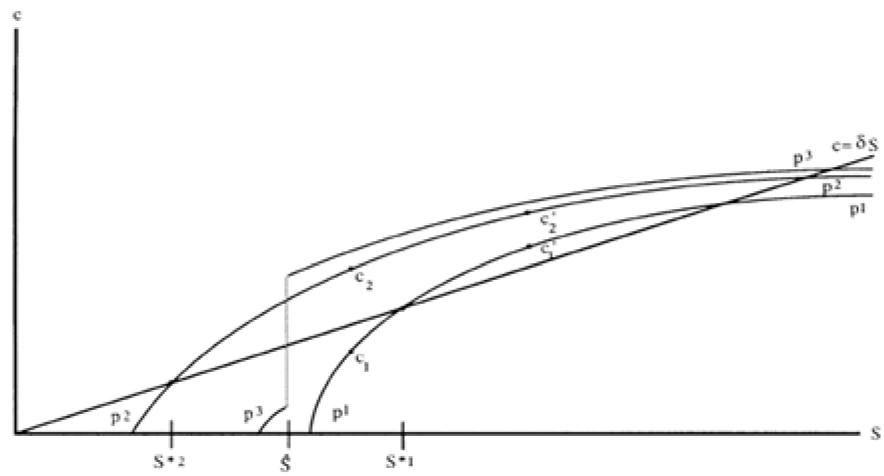



FIG. 2

B&M similarly derive the effect of wealth on steady state consumption by taking the derivative of the FOCs with respect to the marginal utility of wealth ($\mu$):

$$\frac{dc^*}{d\mu} = \frac{\delta}{a_{cc}B}\left[(\sigma + \delta)p_c - \frac{da_S}{d\mu}\right] \tag{13}$$

We can define the addictive good as inferior or superior by the sign $\frac{dc^*}{d\mu} \lesseqgtr 0$. The direction of the inequality depends partly on how your earnings function $w(S)$ is affected by consumption. B&M suggest that harmful goods are more likely to be inferior because the productivity loss associated with addiction is greater for high paying jobs.

## Temporary Price Changes & Life Events

The effects of past and future prices are not symmetric. B&M model this with an additional exponential discount factor in terms of future time $\tau$ (See the last line of A3). At time $t$, the utility discounting on future periods is $e^{-(\sigma+\delta)(\tau-t)}$, compared to the utility discounting of past periods $e^{-(\sigma+\delta)t}$.

The effect of a small, temporary change in price has a non-zero limit as the length of time → 0. B&M give the limits as:

$$\frac{dc(t)}{dp(\tau)} = \frac{(\delta+\lambda_1)e^{-\lambda_2\tau}}{a_{cc}(\lambda_2-\lambda_1)}\left[(\delta + \lambda_2)e^{\lambda_2 t} - (\delta + \lambda_1)e^{\lambda_1 t}\right], \tau > t \tag{14}$$

$$\frac{dc(t)}{dp(\tau)} = \frac{(\delta+\lambda_1)e^{\lambda_1 t}}{a_{cc}(\lambda_2-\lambda_1)}\left[(\delta + \lambda_2)e^{-\lambda_1 \tau} - (\delta + \lambda_1)e^{-\lambda_2 \tau}\right], t > \tau \qquad (15)$$

(14) and (15) imply that the cross-price effects (across time periods) depend on $(\delta + \lambda_1)$. We show in A5 that $(\delta + \lambda_1) > 0$ means that consumption of a good in different time periods are adjacent complements. Therefore, the cross-price effects (14) and (15) are negative when $c(t)$ and $c(t+1)$ are complementary. This responsiveness to future prices is the key difference to distinguish "rational addicts" from "myopic". B&M also suggest that adjacent complements are similar to "complementarity in the Slutsky sense" (somewhat contradicting Ryder and Heal).

The negative cross-price effects associated with complementary goods are greater when the degree of addiction is higher or the anticipation of price changes is longer. This effect is amplified when steady states can be unstable because small changes in price can lead to different steady states with very different levels of consumption (See Fig. 2).

These temporary changes in price effect consumption only in the current period. On the other hand, permanent changes affect the price in all future periods, which leads to further reduction of current period consumption. Thus, the SR price responsiveness is smaller than in the LR.

B&M extend the idea of temporary price changes to temporary life events that stimulate addiction (e.g. divorce, unemployment, death of friend, etc) by increasing the marginal utility of addictive consumption but lowering total utility. B&M address the critics of utility ("happiness")-maximization saying, "Although our model does assume that addicts are rational and maximize utility, they would not be happy if their addiction results from anxiety-raising events, such as death or divorce, that lower their utility. […] However, they would be even more unhappy if they were prevented from consuming addictive goods."

## Cold Turkey & Binges

B&M describe this empirically observed behavior with Fig.2: $c > \delta\hat{S}$ when $S > \hat{S}$, and $c < \delta\hat{S}$ when $S < \hat{S}$. Here, $\hat{S}$ is not a steady-state stock, but it plays a role similar to unstable steady-state stock when the utility function is concave (Figure 1). A slight deviation from $\hat{S}$ leads to drastic change in consumption, either to zero or along the upper part of the $p^s$ curve. This is the "cold turkey" mechanism at work around $\hat{S}$. Once decided the addict will stop consuming immediately.

The explanation of such discontinuity is as follows: If $S$ is even slightly bigger compared to $\hat{S}$, the optimal consumption plan calls for high $c$ in the future because the good is highly addictive. Strong complementarity between present and future consumption then requires a high level of current $c$. If $S$ is even slightly below $\hat{S}$, then future consumption will be very low by strong complementarity. We call $\hat{S}$ the "hooking" point.

## Claims that Addictive Behavior is Consistent with "Rationality"

| Behavior | Explanation | Algebraic presentation | Theoretical and/or Empirical Justification |
|---|---|---|---|
| **Myopia** | Concept of rationality does not rule out strong discount of future events | $\sigma \to \infty$ (if $\sigma = r$) so $a(t) \to 0$ | Fully myopic ($\sigma = \infty$) behavior is formally consistent with the model and, therefore "rational" |
| **Myopia with old age** | Old people are "rationally" myopic and more likely to be addicted | $\frac{1}{t} = \sigma$ – approximation of time preference | Even with high (theoretical) rate of time preference, empirically, they are usually not addicts (B&M attribute to poor health and selection in the type of people who live to old age) |
| **Cold Turkey** | Rational addiction theory implies that highly addictive goods *require* cold turkey to quit because the strong degree of complementarity means the consumption path is discontinuous | $\lambda = \frac{\sigma \pm \sqrt{\sigma^2 + 4B}}{2}$ Both roots are complex, when $4B < -\sigma^2$ | Implies sharp swings in consumption are possible in response to "small" changes in the price |
| **Ending addiction regardless of short-run "pain"** | Rational to exchange a large short-term loss in utility for an even larger long-term gain | $\frac{\partial U}{\partial c} > 0$ implies utility loss from withdrawal; $\frac{\partial U}{\partial s} < 0$ implies long-run gain from subsequently lower $S$ | Addicts are actually forward-looking |
| **Addict's difficulty quitting** | An individual's weak will and limited self-control in their attempt to quit does not imply they are irrational. It may take trial and error to find the best option to reduce the short-run loss in utility from withdrawals. | Similar argument as above | Addicts are utility maximizers and searching for the best way to reduce pains associated with withdrawal. |
| **Overeating** | This is *not* a prototype of irrational behavior. | Suppose a weight stock ($S_w$), eating capital stock ($S_e$), and rates of depreciation ($\delta_w$, $\delta_e$): $\alpha_{cs_w} < 0$ , $\alpha_{cs_e} > 0$ ; $\delta_e > \delta_w$, $\Rightarrow \dot{S_e} > \dot{S_w}$ | Cycles of overeating and dieting possible: After increase in $S_e$, eating levels off and begins to fall so that weight stock falls. When weight is sufficiently low, eating picks up the cycle starts again. |
| **Binges** | Just like overeating, binges are not an inconsistent behavior. | Could think of similar stocks related to drinking: "social stock" and "productivity stock" | Similar to overeating case: cycles can be the outcome of consistent maximization over time, recognizing that the effect of increased current drinking/socializing, on both future productivity and the desire to drink more in the future. |

## Empirical Evidence

Within the paper B&M cite some empirical evidence to give credence to their model. For example, they include empirical measures of price elasticity for cigarettes. They also cite long-term changes in smoking behavior after the full-price of cigarettes changed with new health warnings from the Surgeon General. Since B&M, the literature of economic theory and empirical microeconomic research on addiction has greatly expanded. The economic framework for modeling addictive behaviors has contributed to the cross- and inter-disciplinary subject of addiction. “Rational Addiction” now has more than 3000 citations in Google Scholar.

The theoretical model of rational addiction “has been reinforced by a sizable empirical literature, […] which has tested and generally supported the key empirical contention of the Becker and Murphy model: that the consumption of addictive goods today will depend not only on past consumption but on future consumption as well. More specifically, this literature has generally assessed whether higher prices next year lead to lower consumption today, as would be expected with forward-looking addicts. The fairly consistent findings across a variety of papers that this is the case has led to the acceptance of this framework for modeling addiction.” (Gruber and Köszegi, 2001).

Becker, Grossman and Murphy (1994) were the among the first to empirically test “rational addiction” with cigarette prices and consumption. They observed a negative cross-price effect of future prices on current consumption of cigarettes, and found greater long-run price elasticity compared to the short-run.

Similarly, Gruber and Köszegi (2001) found empirical evidence of forward-looking cigarette smokers. They estimated the consumption response to state excise taxes that have been legislated but not yet implemented. They also extend the theoretical model of B&M with hyperbolic discounting, which delivers the same theoretical predictions as the rational addiction model. Empirically, most of the tests of rational addiction are not robust to the possibility of hyperbolic discounting.

Wang (2014) modeled smoking behavior as a dynamic discrete choice model, where smokers decide whether to smoke or quit in each period based on their health, income, survival probability, and state transition functions. They estimate the discount factor with maximum likelihood and find β close to 1, which indicates that smokers are forward-looking in their behavior.

## Criticism

### Hyperbolic Discounting

B&M assume *time consistent* discounting, and this assumption generated much criticism. The basis of this criticism comes from the overwhelming psychological evidence that humans are time-inconsistent. This means that the relative valuation of utilities in the near future is different compared to the relative valuation of utilities in the far future. George Ainslie’s works, *Picoeconomics* (1992) and *Breakdown of Will* (2001) are often cited in this regard. However, Gruber and Koszegi (2001) extend the rational addiction model with hyperbolic discounting and find that many of the implications are the same. However, optimal taxation of harmful addictive goods does differ if addicts are actually hyperbolic discounters.

### “The Dangers in Analogical Thinking” & Misapplied Math

Other critics argue that “rational” modeling of addiction is an absurd approach and inappropriate application of math. Elster (1997) is a strong critic of rational addiction theory. B&M model increasing rate of time discounting and assume that an addict’s “*awareness* of the future consequences is not impaired.” Elster maintains that addicts often cannot make rational, utility-maximizing decisions about the future because addictions (e.g. mind-altering drugs) impair judgment.

Rogeberg (2004) also takes a stab at B&M’s claim of rational addiction behavior. He argues that the application of “rational thinking” economic theory to addiction is “absurd,” and that this is an inappropriate use of math to defend a theory. Similar to Elster, the main assumption of contention for Rogeberg is that addicts make detailed and forward-looking plans.

We think that these types of criticism may qualify B&M’s theory for certain addictive substances that seriously affect judgment (e.g some types of illicit drugs). However, the case can be made that “Rational Addiction” applies to many other addictions that do not affect decision-making, although they may have real physiological effects. For example, you can become addicted to a caffeine buzz, but coffee does not affect your ability to make rational, forward-thinking decisions. In fact, most coffee-drinkers think that caffeine improves their brain function and focus!

## Extensions

For simplicity in the dynamic analysis, B&M assume that $h[D(t)] = 0$, but we think that $D(t)$ might be a choice variable of interest. B&M call $D(t)$ “expenditures on endogenous depreciation or appreciation.” Then, $h[D(t)]$ enters the state transition function $\dot{S}(t) = c(t) - \delta S(t) - h[D(t)]$. If your consumption stock of addictive goods comes from “learning” from past consumption, we can think of $D(t)$ as expenditures that help you “forget” your past consumption. For example, you might spend money on rehabilitation or new hobbies in order to reduce the effect of your stock on future consumption. We will presume that all expenditures $D$ fight addiction that $h$ is a monotonic function of $D$.

With $D(t)$ remaining in the model, the Taylor Series Expansion from (6) contains terms for the partial effects of $D$ through $S$, as well as the linear term for the price of $D$:

$$F(t) = \alpha_c c(t) + \alpha_S S(t) + \frac{\alpha_{cc}}{2}[c(t)]^2 + \frac{\alpha_{SS}}{2}[S(t)]^2 + \alpha_{cS}c(t)S(t) + \boldsymbol{\alpha_{SD}S(t)D(t)} - \mu p_c c(t) - \boldsymbol{\mu p_D D(t)} \qquad (16)$$

Now, we suggest the condition for adjacent complementarity becomes:

$$(\sigma + 2\delta)\alpha_{cS}' > -(\alpha_{SS} + \alpha_{SD}) > 0 \qquad (17)$$

As before, $\alpha_{SS} < 0$, and now we assume $\alpha_{SD} < 0$. Thus, in order for the adjacent complementarity to hold with additional anti-addiction measures, $\alpha_{cS}'$ must be very large. This makes intuitive sense, that a strong addiction would require more resources to defeat. More interesting questions involve the optimal path of $D(t)$ and how it relates to different types of addiction (with different degrees of $\alpha_{cS}'$). One could solve the dynamic optimization problem with the value function (16) to derive the optimal path of $c(t)$, $y(t)$, $S(t)$, and $D(t)$.

## Summary & Conclusions

Becker and Murphy develop a model of addiction that hinges on the relationship between past and present consumption of an addictive good. Consumption in different periods is related by a “stock” variable $S$. They derive the optimal paths of “consumption stock” $S(t)$ and consumption $c(t)$, and then compare the dynamics of $c$ and $S$ near the steady-state. They find this dynamic model of forward-thinking addicts is consistent with many stylized facts and empirically-testable behaviors of addiction, such as price responsiveness, the increased consumption with stressful life events, binges, purges, and cold-turkey quitting. The body of empirical work, mostly related to smoking and drinking, gives strong support for B&M’s rational addiction model.

## References


Barro, Robert J. and Xavier Sala-i-Martin. 2004. *Economic Growth*, Second Edition. Cambridge, MA: The MIT Press.

Becker, Gary S. and Kevin M. Murphy. 1988. A theory of rational addiction. *Journal of Political Economy* 96(4): 675-700.

Becker, Gary S., Michael Grossman, and Kevin M. Murphy. 2994. An empirical analysis of cigarette addiction. *The American Economic Review.* 84(3):396-418.

Elster, Jon. 1997. More than enough. *The University of Chicago Law Review, 64*(2), 749-764.

Gruber, Jonathan and Botond Köszegi. 2001. Is addiction “rational”? Theory and evidence. *The Quarterly Journal of Economics* 116(4): 1261-1303.

Rogeberg, Ole. 2004. Taking absurd theories seriously: Economics and the case of rational addiction theories. *Philosophy of Science, 71*(3), 263-285. < http://www.jstor.org/stable/10.1086/421535 >.

Qin, Wei. 2012. Hyperbolic discounting. < http://tinyurl.com/p2fp9qg >.

Wang, Yang. 2014. Dynamic implications of subjective expectations: Evidence from adult smokers. *American Economic Journal: Applied Econometrics* 6(1): 1-37.


# Appendix

### A1. Time Preference

Individuals discount utility from future periods, and there is empirical evidence of different types and magnitudes of discounting behavior. In essence, time preference reflects a person’s lack of foresight or self-control. Theoretical models with multiple periods (or continuous time) assume some form of discounting which can have implications for the results of the model.

B&M assume exponential discounting with a constant rate of time preference ($\sigma$), which enters the model through $e^{-\sigma t}$. This is a type of time-consistent preference, which is the type typically used in rational choice theory. It implies that the marginal utility of substitution between goods at any two points in time depends only on how far apart those two points are, and the valuation of utility falls at constant rate over time.

Non-constant time preference, on the other hand, creates a time-consistency problem. The time-inconsistency arises when the relative valuation of utilities at different time changes with the time horizon. One is much more impatient about consumption between today and tomorrow, than between future consumption in one month and one month and a day. Hyperbolic discounting is one kind of time-inconsistent preference, and the form of the discounting factor is $(1+\alpha t)^{\beta}$ with time $t$ and parameters $\alpha$ and $\beta$. (See Qin 2012).

## A2. Control Problem & First Order Conditions

We will use a Hamiltonian[1] to solve the dynamic optimization problem (i.e. control problem). The control variables are $y$, $c$, and $D$ and the state variable is $S$. The function $\dot{S}(t) = c(t) - \delta S(t) - h[D(t)]$ describes the transition in $S$. The problem is setup as the intermediate function with multipliers $\boldsymbol{\mu} = (\mu_1, \mu_2)$ for the two constraints, the state transition function $\dot{S}(t)$ and the budget constraint.

$$H = e^{-\sigma t}u[y(t), c(t), s(t)] + a(t)\big[c(t) - \delta S(t) - h[D(t)]\big] - \mu[e^{-rt}[y(t) + p_c(t)c(t) + p_d(t)D(t)] - A_0 - e^{-rt}w\big(S(t)\big)]$$

To derive the FOCs, take the derivative of $H$ w.r.t. the state variables $y$, $c$, and $D$:

$$\frac{\partial H}{\partial y} = e^{-\sigma t}\frac{\partial u}{\partial y} - \mu e^{-rt} = 0 \quad \Longrightarrow \quad u_y = \mu e^{(\sigma - r)t}$$

$$\frac{\partial H}{\partial c} = e^{-\sigma t}\frac{\partial u}{\partial c} + a(t) - \mu e^{-rt}p_c(t) = 0 \quad \Longrightarrow \quad u_c = \mu e^{-rt}p_c(t)e^{\sigma t} - e^{\sigma t}a(t) = \mu p_c(t)e^{(\sigma - r)t} - e^{\sigma t}a(t)$$

$$\frac{\partial H}{\partial D} = a(t)\frac{\partial h}{\partial D} - \mu e^{-rt}p_d(t) = 0 \quad \Longrightarrow \quad h_d \cdot a(t) = \mu e^{-rt}p_d(t)$$

B&M denote the multipliers $(\mu_1, \mu_2) = (a(t), \mu_2)$ where $a(t)$ is the shadow price of consumption stock ($S$). We can derive $a(t)$ by taking the derivative of the Hamiltonian with respect to $S$ and setting it equal to the negative of the change in the co-state variable.

$$\frac{\partial H}{\partial S} = e^{-\sigma t}\frac{\partial u}{\partial S} - \mu_1\delta + \mu_2 e^{-rt}\frac{\partial w}{\partial S} \quad \Longrightarrow \quad e^{-\sigma t}\frac{\partial u}{\partial S} - a(t)\delta + \mu e^{-rt}\frac{\partial w}{\partial S} = -\dot{a}(t)$$

To analytically solve this differential equation, we multiply both sides by an integration factor $e^{-\delta t}$ where $\delta$ is the coefficient on the co-state variable from the derivative $\frac{\partial H}{\partial S}$.

$$e^{-(\sigma+\delta)t}\frac{\partial u}{\partial S} - e^{-\delta t}a(t)\delta + \mu e^{-(r+\delta)t}\frac{\partial w}{\partial S} = -e^{-\delta t}\dot{a}(t)$$

$$\Longrightarrow \quad -e^{-\delta t}[\dot{a}(t) - a(t)\delta] = e^{-(\sigma+\delta)t}u_s + \mu e^{-(r+\delta)t}w_s$$

[1] The mathematical appendix A.3 in *Economic Growth* by Barro and Sala-i-Martin was helpful to set up the Hamiltonian for B&M model and derive the FOCs.

$$\Longrightarrow \; -\frac{d}{dt}\left[e^{-\delta t}a(t)\right] = e^{-(\sigma+\delta)t}u_s + \mu e^{-(r+\delta)t}w_s$$

Integrating both sides,

$$-\int \frac{d}{dt}\left[e^{-\delta t}a(t)\right]dt = \int e^{-(\sigma+\delta)t}u_s + \mu e^{-(r+\delta)t}w_s dt$$

$$\Longrightarrow \quad a(t) = \int e^{-(\sigma+\delta)t}u_s dt + \int \mu e^{-(r+\delta)t}w_s dt$$

If we pick a fixed $t\epsilon[0,T]$, then $a(t) = \int_t^T e^{-(\sigma+\delta)(\tau-t)}u_s d\tau + \int_t^T \mu e^{-(r+\delta)(\tau-t)}w_s d\tau$

## A3. Euler's Equation

$$e^{-\sigma t}F[S(t),c(t)] = e^{-\sigma t}\left[\alpha_c c(t) + \alpha_S S(t) + \frac{\alpha_{cc}}{2}[c(t)]^2 + \frac{\alpha_{SS}}{2}[S(t)]^2 + \alpha_{cS}c(t)S(t) - \mu p_c c(t)\right]$$

We know that $c(t) = \dot{S}(t) + \delta S(t)$ from the state transition function, so we can write the intermediate function in terms of only $S(t)$ and $\dot{S}(t)$:

$$e^{-\sigma t}F[S(t),c(t)] = e^{-\sigma t}\Big[\alpha_c\left(\dot{S}(t) + \delta S(t)\right) + \alpha_S S(t) + \frac{\alpha_{cc}}{2}\left(\dot{S}(t) + \delta S(t)\right)^2 + \frac{\alpha_{SS}}{2}[S(t)]^2 +$$
$$\alpha_{cS}\left(\dot{S}(t) + \delta S(t)\right)S(t) - \mu p_c\left(\dot{S}(t) + \delta S(t)\right)\Big]$$

Rearranging on the RHS, we get the following as the intermediate function:

$$I = e^{-\sigma t}\Big[\dot{S}(\alpha_c - \mu p_c) + S(\alpha_S + \alpha_c\delta - \mu p_c\delta) + \dot{S}\cdot S(\alpha_{cc}\delta + \alpha_{cS}) + S^2\left(\frac{\alpha_{cc}}{2}\delta^2 + \frac{\alpha_{SS}}{2} + \alpha_{cS}\delta\right) +$$
$$\dot{S}^2\frac{\alpha_{cc}}{2}\Big]$$

Then calculate the Euler equation: $\frac{\partial I}{\partial S} - \frac{d}{dt}\left(\frac{\partial I}{\partial \dot{S}}\right) = 0$.

$$\frac{\partial I}{\partial S} = e^{-\sigma t}\left[(\alpha_S + \alpha_c\delta - \mu p_c\delta) + \dot{S}(\alpha_{cc}\delta + \alpha_{cS}) + S(\alpha_{cc}\delta^2 + \alpha_{SS} + 2\alpha_{cS}\delta)\right]$$

$$\frac{\partial I}{\partial \dot{S}} = e^{-\sigma t}\left[(\alpha_c - \mu p_c) + S(\alpha_{cc}\delta + \alpha_{cS}) + \dot{S}\alpha_{cc}\right]$$

$$\frac{d}{dt}\left(\frac{\partial I}{\partial \dot{S}}\right) = e^{-\sigma t}\left[\dot{S}(\alpha_{cc}\delta + \alpha_{cS}) + \ddot{S}\alpha_{cc}\right] - \sigma e^{-\sigma t}\left[(\alpha_c - \mu p_c) + S(\alpha_{cc}\delta + \alpha_{cS}) + \dot{S}\alpha_{cc}\right]$$

$$\frac{\partial I}{\partial S} - \frac{d}{dt}\left(\frac{\partial I}{\partial \dot{S}}\right) =$$
$$e^{-\sigma t}\left[(\alpha_S + \alpha_c\delta - \mu p_c\delta) + \dot{S}(\alpha_{cc}\delta + \alpha_{cS}) + S(\alpha_{cc}\delta^2 + \alpha_{SS} + 2\alpha_{cS}\delta)\right] - e^{-\sigma t}\left[\dot{S}(\alpha_{cc}\delta + \alpha_{cS}) +\right.$$
$$\left.\ddot{S}\alpha_{cc}\right] + \sigma e^{-\sigma t}\left[(\alpha_c - \mu p_c) + S(\alpha_{cc}\delta + \alpha_{cS}) + \dot{S}\alpha_{cc}\right] = 0$$

$$\Longrightarrow \quad \ddot{S} - \sigma\dot{S} - S\left(\sigma\delta + \delta^2 + \frac{\alpha_{SS}}{\alpha_{cc}} + \sigma\frac{\alpha_{cS}}{\alpha_{cc}} + 2\delta\frac{\alpha_{cS}}{\alpha_{cc}}\right) = \frac{(\sigma+\delta)\alpha_c + \alpha_S - (\sigma+\delta)\,\mu p_c}{\alpha_{cc}}$$

$$\Longrightarrow \quad \ddot{S} - \sigma\dot{S} - BS = \frac{(\sigma+\delta)\alpha_c + \alpha_S}{\alpha_{cc}} - \frac{(\sigma+\delta)\mu p_c}{\alpha_{cc}}\text{; with } B = \delta(\sigma+\delta) + \frac{\alpha_{SS}}{\alpha_{cc}} + (\sigma + 2\delta)\,\frac{\alpha_{cS}}{\alpha_{cc}}$$

## A4. Analytical Solution of the ODE

The solution to the differential equation (9) is of the form:

$$\frac{-b\pm\sqrt{b^2-4ac}}{2a} = \frac{\sigma\pm\sqrt{\sigma^2-4B}}{2} \tag{A4.1}$$

The term under the radical is positive because it can basically be written as a quadratic function:

$\sigma^2 - 4B = \frac{1}{a_{cc}}[(\sigma+2\delta)^2 a_{cc} + 4a_{ss} + 4(\sigma+2\delta)a_{cs}] \approx (a+b)^2 = [(\sigma+2\delta)+2]^2$ , so that the positive terms $(\sigma+2\delta)^2$ and $4\frac{a_{ss}}{a_{cc}}$ outweigh the potentially negative term $4(\sigma+2\delta)\frac{a_{cs}}{a_{cc}}$ and $\sigma^2 - 4B > 0$. Therefore, the two roots (A4.1) are real. We also know that the roots define a consumption path that *maximizes* utility because $F$ is strongly concave in $c$ and $S$ and the Hessian matrix is negative definite.

The optimal path of $S$ comes from the solution to the differential equation: $S(t) = de^{\lambda_1 t} + ce^{\lambda_2 t}$, with some constants $c$ and $d$. We do not consider the second root $\lambda_2 = \frac{\sigma+\sqrt{\sigma^2-4B}}{2} > \frac{\sigma}{2}$ because we cannot impose a transversality condition to bound $[c(t)]^2$ in the value function (7) because:

$$\lim_{t\to\infty} e^{-\sigma t}e^{\lambda_2 t}\,[c(t)]^2 = \lim_{t\to\infty} e^{(\lambda_2-\sigma)t}\,[c(t)]^2 = \infty$$

Therefore, the optimal path of $S$ is determined by the smaller root ($\lambda_1$) and the initial stock ($S_0$):

$$S(t) = de^{\lambda_1 t} + S^* \text{ , where } d = S_0 - S^* \tag{A4.2}$$

$$\Rightarrow \quad S(t) = e^{\lambda_1 t}S_0 - e^{\lambda_1 t}S^* + S^*$$

$$\Rightarrow \quad \dot{S}(t) = \lambda_1 e^{\lambda_1 t}S_0 - \lambda_1 e^{\lambda_1 t}S^*$$

Then, we can also find the optimal path of $c(t)$:

$$c(t) = \dot{S}(t) + \delta S(t) = (\delta+\lambda_1)S(t) - \lambda_1 S^* \tag{A4.3}$$

The steady state $S^*$ is stable only if $B > 0$ so that $\lambda_1 = \frac{\sigma-\sqrt{\sigma^2-4B}}{2} < 0$.

## A5. Condition for Adjacent Complementarity

From (A4.3), the relationship between $c$ and $S$ is defined by the sign of $(\delta+\lambda_1)$.

$$\delta+\lambda_1 = \delta + \frac{\sigma}{2} - \frac{\sqrt{\frac{1}{a_{cc}}\left[(\sigma+2\delta)^2 a_{cc} + 4a_{ss} + 4(\sigma+2\delta)a_{cs}\right]}}{2}$$

$$\delta+\lambda_1 = \frac{(\sigma+2\delta)}{2} - \frac{\sqrt{(\sigma+2\delta)^2 + \frac{4a_{ss}+4(\sigma+2\delta)a_{cs}}{a_{cc}}}}{2}$$

Since $a_{ss} < 0$ and $a_{cc} < 0$, then the sign of $\delta + \lambda_1$ is determined by the sign on the term $\frac{4a_{ss}+4(\sigma+2\delta)a_{cs}}{a_{cc}}$. It follows that sign of $(\delta + \lambda_1)$ and hence relationship of $S$ and $c$ is determined by the following:

$$(\sigma + 2\delta)\alpha_{cs} \gtreqless -\alpha_{ss} > 0 \tag{A5.1}$$

(A5.1) Defines adjacent complementarity: if LHS is greater than RHS, then $\delta + \lambda_1 > 0$, so $c$ and $S$ are positively related.