CONTEMPLATION VS. INTUITION. A REINFORCEMENT LEARNING PERSPECTIVE.

IN-KOO CHO[♯] AND ANNA RUBINCHIK[♮]

Abstract. In a search for a positive model of decision-making with observable primitives, we rely on the burgeoning literature in cognitive neuroscience to construct a three-element machine (agent). Its control unit initiates either impulsive or cognitive elements to solve a problem in a stationary Markov environment, the element "chosen" depends on whether the problem is mundane or novel, memory of past successes and the strength of inhibition.

Our predictions are based on a stationary asymptotic distribution of the memory, which, depending on the parameters, can generate different "characters", e.g., an *uptight dimwit*, who could succeed more often with less inhibition, as well as a *laid-back wise-guy*, who could gain more with a stronger inhibition of impulsive (intuitive) responses.

As one would expect, stronger inhibition and lower cognitive costs increase the frequency of decisions made by the cognitive element. More surprisingly, increasing the "carrot" and reducing the "stick" (being in a more supportive environment) enhances contemplative decisions (made by the cognitive unit) for an alert agent, i.e., the one who identifies novel problems frequently enough.

J.E.L. *Classification numbers.* D01.
Keywords: the two-system decision-making, executive control, inhibition, adaptive learning, stochastic approximation.

We would like to thank for its hospitality the Department of Economics at Tel-Aviv University and the Department of Economics at the University of Queensland, Australia. We thank for their comments and suggestions Sh. benYosef, I. Erev, I. Gilboa, A. Goldenshluger, J. Kline, A. McLennan, R. Smorodinsky, as well as the participants of *The International Workshop on Decision Making, Game Theory and Epistemic Logic* at the University of Queensland and the participants of the departmental seminar at the University of Haifa.

The scientific responsibility is assumed by the authors.

[♯]Department of Economics, University of Illinois, USA. inkoocho@uiuc.edu.

[♮]Department of Economics, University of Haifa, Israel.
arubinchik@econ.haifa.ac.il.

1. Conceptual motivation and related research

Our goal is to describe a technology, the technology of decision making. Why? Even the most ardent supporters of rational decision-making may be disappointed with their own decisions that are made hastily, without much thought. Indeed, how often do we know what *should be* done, but act differently? More importantly, the observed individual behavior departs from rationality (even most broadly defined as in Gilboa and Schmeidler (2001)), and this is a well-established fact.[1] We do not intend to enter a debate about economic significance of this departure, however, we hope the reader will agree that it is useful to construct a model that could help us understand when we are more likely to buy what we do not need or trust unreliable politicians, at least for building meaningful predictions.

Also, we are not even attempting to offer a glimpse into the recent literature that uses an axiomatic approach to incorporate various forms of "irrational" behavior, though, no doubt, it produced most elegant and insightful results.[2] To stress, in no way are we disputing the importance of the traditional primitives for clear and logical analysis of how decisions *should* be made. Our goal is to use "natural laws" to describe how decisions *are* being made. The ability to depict individual behavior in a more realistic way, at least identifying its basic elements, makes the previously puzzling "irrational behavior" quite "normal" and provides useful insights into creating better-tuned motivation schemes for traditional economic settings, just as a more realistic description of production technology can help in applying the classical economic analysis properly.

Here is our approach in a nutshell. "Nature" has endowed individuals with a machine which can identify problems and produce solutions (successful or not). We wish to look at this machine a little more closely. Recall that within a classical approach the process of making decisions is taken as a black box and the box is replaced by a rational calculation, in a hope that it imitates the actual behavior. Our main concern with this approach is that it overlooks the trivial "shortcuts" that are used by nature, especially because resorting to such shortcuts depends on individual experience and traits. It is those shortcuts that generate, at times, what might seem as "silly mistakes" to a rational observer. To be more precise we rely on selected findings from biology

[1] Cf. the work of Daniel Kahneman and Amos Tversky, or of Malcolm Gladwell for example.

[2] A.o., it includes research in neuroeconomics, see Glimcher and Fehr (2013) for an overview.

and cognitive psychology and sketch briefly the closely related literature in economics.[3]

1.1. Stylized facts about the "technology" of decision-making.

1.1.1. *Thinking requires resources.* Recent studies (McNay et al., 2000) show that glucose is depleted faster after more difficult tasks and in those areas in the brain of rats which has been used for task completion.

The idea that cognitive activity requires resources that have an alternative cost, of course, has been used even in economics, see for example, Woodford (2012) and Sims et al. (2010), who provide a rationale for various degrees of sophistication in solving problems. One has to keep in mind though that deliberately choosing the amount of resources to dedicate to thinking might appear to be a more complicated problem than the original one, thus, potentially, requiring more resources than the amount to be allocated in the first place. Hence such a decision can not involve elaborate "forward-looking" calculations, rather, it has to be fully based on past experience.

In addition, there is some evidence that fast and easy actions trigger immediate rewards,[4] and that can be considered as an alternative cost of thinking. Yet another component of the cost is the foregone time: decisions that require thinking are typically slower, cf. e.g., Kahneman (2011).

It is clear that many decisions that we have to make are rather mundane, starting from the sequence of moves that are needed to eat, walk, breathe and, finally, to control the heart-beat. Not surprisingly, most of the time we are not even aware of making such decisions, for if we were, the cost would have been prohibitively high. Perhaps, that is why in order to survive, we have to resort, at least sometimes, to a simpler way of choosing how to act.

1.1.2. *Existence of different decision-making units.* To start, let us mention the pioneering work of Ferrier (1876) who distinguished between impulsive and deliberate modes of decision making:

> If the centres of inhibition, and thereby the faculty of attention, are weak, or present impulses unusually strong, volition is impulsive rather than deliberate. (Ferrier, 1876, p. 287).

[3]For an overview in the latter field, see, for example, Corr and Matthews (2009, parts IV, V), in particular, uncovering physiological explanations for various "personality traits" and their effect on behavior.

[4]See Chung et al. (2015).

As for "state of the art" in neurobiology on the subject, cf. the review by Bari and Robbins (2013), who stress the hierarchical nature of the process and discuss the role of the control center:

> In general, executive control is thought to operate in a hierarchical manner with the PFC [prefrontal cortex] having a leading role over lower-level structures...(Bari and Robbins, 2013).

The 'machine' we model has three basic components: a *cognitive* unit, an *impulsive* unit and a *control* unit. Thus, we distinguish between pure "problem solving" tasks that might involve activating working and long-term memory, done by our *cognitive unit*, and executive control function (the *control* unit), which is responsible for inhibiting fast "instinctive" responses (by the *impulsive* unit), a.o., in go/no-go tasks. The related neurobiological research is burgeoning,[5] although there are still a lot of open questions. Nevertheless, being able to identify the part of the brain responsible for a particular impulse inhibition allows not only to detect its activation during various tasks, but also to carefully study the relevant triggers which alter the resulting behaviour (Spieser et al., 2015). With the triggers weakening or enhancing the control unit, the picture of the brain operation emerging from this literature is, indeed, that of a 'machine'.

Not surprisingly, the distinction between "fast" and "slow" decision-making has found its way into economics literature, probably due to the recent book by Kahneman (2011). In fact, a multi-level decision process appeared already in the celebrated chain-store paradox by Selten (1978). The recent two-mode decision making models, Alos Ferrer (2013); Cerigioni (2015), still retain, at least in part, the classical primitives (preferences), which we do not adopt here.

One might also connect the literature on multiple selves to this discussion, however such models typically involve deliberately reasoning players who are able to compute their equilibrium strategies, with some of the players being allowed to entertain "naive" (inconsistent) beliefs, cf. e.g., Spiegler (2011). The multi-selves dichotomy is different from ours: we are not aiming at describing the decision-making process in different modes, frames of mind or periods of time. The cognitive and impulsive units in our model operate simultaneously at the time when the decision is to be made. The eventual action can be triggered by either, how to channel the decision is subject to the executive unit.

[5]Cf. Smith and Jonides (1999); Shomstein (2012); Bari and Robbins (2013); Ding et al. (2014).

The units are not strategic players, they operate according to "laws of nature" taking past experience as an input.

1.1.3. *Interpersonal differences in inhibition control.* These differences might be responsible for individual traits (some forms of impulsiveness and even risk-taking),[6] which affect one's behavior sometimes against one's own will.[7] In part, we will attempt to identify the strength of the inhibition control (parameter h in our model) from the observed behavior, i.e., activation of cognitive or impulsive unit to solve problems, and that we take as an "observed variable".

1.1.4. *Fast and easy actions trigger immediate rewards.* The three decision-making units are analyzed in the context-dependent choice environment in the recent study by Chung et al. (2015). Among other findings is the observation that fast decision-making bears a pleasurable reward immediately following the impulsive act, as was already mentioned above. This might be due to conditioning, see e.g., Schultz et al. (1997) for the description of the underlying mechanism.

1.1.5. *Experience matters.* Past experiences are encoded in memory. Here we focus on "emotional" memory of past successes and failures, see, e.g., Fujiwara and Markowitsch (2006) for a summary of the relevant studies.

1.1.6. *Novel situations might activate cognitive functions.* For partial support see Corbetta et al. (2008): the nature of the mechanism is rather involved. A casual observation suggests that being startled (by a puzzle), one is likely to start thinking (and exploring) provided the situation is not life-threatening.

We proceed with a formal description of the decision making process intertwined with interpretations which, in part, justify the assumptions.

2. The model

2.1. **The decision-making process.** As mentioned in the introduction, the decision-making process has three components (units): impulsive, cognitive, and control. The cognitive (or contemplative) unit expands resources during its search for an appropriate action. The impulsive unit is fast, provides an answer based on "intuition" and requires minuscule resources to operate.

[6] Although some of these connections are subject to additional scrutiny, Brown et al. (2015).

[7] Cf. von Hippel and Dunlop (2005), see also an overview in Kahneman (2011).

Interpretation. One can think of an habitual response that was learned in the past and that has been dopamine-encoded (Kim et al., 2015) in the long-term memory or an instinctive reaction as examples of actions driven by the impulsive unit.

Control unit can stop the impulsive one from triggering the action and shift the decision to the cognitive unit. Exercising control has a cost ($\kappa > 0$): first, the foregone pleasure from the impulsive decision; second, the added reaction time; and third, the energy spent by operating the cognitive unit itself.

2.2. **The environment.** Here we abstract away from the process of finding a solution to every problem, taking it as a black box, acknowledging however, that many interesting issues are omitted by doing so. We start with a rather informal description of the environment.

There are 2 possible problems in the environment faced by our decision maker (DM): simple and difficult (novel).

Each period DM is confronted with a simple problem with probability $0 < w < 1$, and the difficult one shows up with the complementary probability.

Interpretation. w measures how mundane/challenging is the environment surrounding the DM. In general one might want to have w potentially changing with experience, as in the same environment difficult problems become simple with skill acquisition, but our DM constantly faces new challenges with probability $1 - w$.

A simple problem is similar to the problems encountered in the past, and the impulsive unit identifies the right solution, i.e., triggers an act that generates a reward, with probability $0 < p < 1$ each time and cognitive unit — with probability $0 < a \leq 1$.

Interpretation. Although one might think that in simple situations the learned ("automatic") response should be the right one, we still allow for a possibility of a mistake, thus $p < 1$: impulsive unit might "overreact" to a particular stimulus (signal) and hence trigger an erroneous action. $a \leq 1$ embeds the possibility of failing to find solutions for simple problems, even when thinking is involved.

The control unit channels the decision making based on the history of past successes and the type of task.[8] Either it lets the impulsive unit trigger the action or inhibits the immediate response and shifts the control to the cognitive unit.

A difficult or a novel situation requires introspection and is identified as such (by the control) with probability $0 < z < 1$. If it is identified as difficult, the control activates the cognitive unit, which succeeds in finding the solution (and being rewarded) with probability $0 < g < 1$, it incurs a loss in case of failure. If the problem is not identified as difficult, then the same algorithm is used as in the case of a simple problem: the most successful unit of the two becomes responsible for the decision. If impulsive unit is activated to solve the novel problem, the success rate is $0 < b < 1$.

Interpretation. Some new situations are obviously different, others have more hidden signs, thus are harder to detect, so $z < 1$. There is no clear way to teach to identify new situations, so z probably reflects an in-born quality. Missing relevant information or cognitive limitations in general are reflected in assuming $g < 1$. In addition, some failures are due to experimentation, an inherent feature of learning, which might be triggered upon encounter with a novel problem. Finally, $b < 1$, reflects the probability of making "silly mistakes" when acting instinctively based on a single trigger (thus omitting relevant inputs) in a novel environment.

Note that we did not assume that the cognitive unit always outperforms the impulsive one in finding the right solutions to various problems: in fact, there are many decisions for which the opposite might be true (e.g., choosing how to breathe, walk and occasionally, how to drive a vehicle). One might conjecture that if the cognitive unit is less successful and more costly, it should not be used at all, unless called for by the novelty of the problem (which we assumed), but, surprisingly, this conjecture might be wrong. However, as will be evident once the description of the model is complete (in the next subsection), there is no single unit that is responsible for optimal resource allocation, so one should not expect the system to behave optimally in any way.[9]

[8]The exact mechanism is not fully understood yet, cf. Boureau et al. (2015) for the most recent review.

[9]One could, of course, evoke an argument involving evolutionary pressures to claim that wasteful resource allocation should not be long-lived, but this is beyond the scope of the current investigation.

2.3. **Memory and control unit "decisions".** Memory at time t is a pair of real numbers, I_t, C_t:

- I_t indicates (relative) performance of the impulsive unit, i.e., observed frequency of its successes;
- C_t, similarly, reflects the performance of the cognitive component less the cost, $\kappa > 0$, expanded from its activation.

Given the inhibition parameter $h \in \mathbb{R}$, if a simple problem is encountered control unit activates

cognitive unit if	$I_t < h + C_t$
impulsive unit if	$I_t > h + C_t$
either one	otherwise

If a novel/difficult problem is encountered, the algorithm is the same, unless the novelty is detected and then the cognitive unit is used, independently of the memory, I_t and C_t.

Interpretation. Following section 1.1.3 in the introduction, we view the inhibition parameter, h, mainly as an in-born trait, although one might entertain a view that at least partially, the strength of inhibition can be modified by the appropriate training. To stress, again, although this parameter can not be readily "measured", which unit is activated — impulsive or cognitive — can, by and large, be observed. The inhibition parameter is not necessarily positive, $h < 0$ could stand for inducing an intuitive response, even when the impulses are not very strong.

An action generates a reward if successful, and triggers a loss in case of failure. r and l stand for reward and loss, correspondingly, each depending on whether the problem is simple, m, or difficult, f. This "instantaneous" payoff reinforces the activated unit, i.e., is used to update the indicators of the relative performance.

Interpretation. We assume that the payoffs include the "immediate" gratification from solving the problem correctly in the first case and, similarly, discouragement in case of failure. These parameters can be thought of as reflecting the environment (nurture): more approval for success (from peers, parents, others) translates into higher r and more discouragement, heavier punishment yields higher l.

Of course, these rewards and losses are also, in part, a reflection of the subject's own attitude: how critical the DM is of his failures and how proud he is of success. We could potentially connect it to a

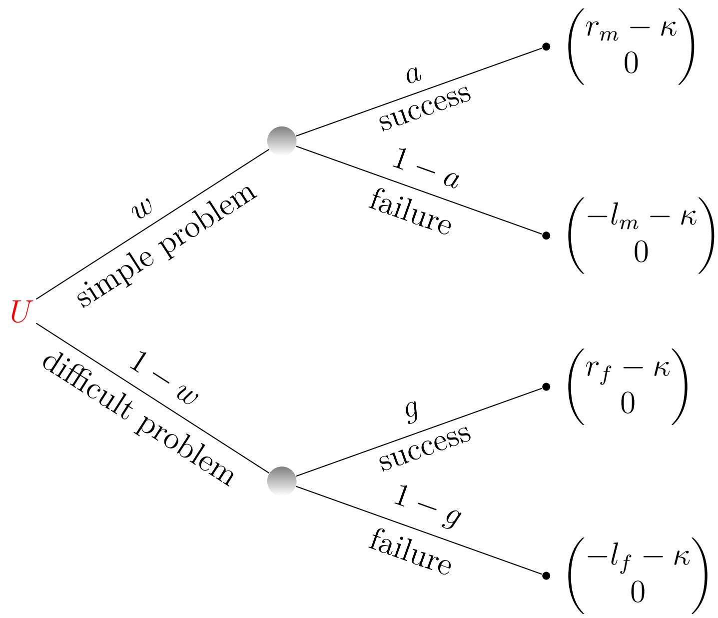

FIGURE 2.1. Random payoff, U, in the cognitive regime. Notice that the impulsive unit is not rewarded, so the second component of the realized payoff in U is zero.

variable of choice: one can decide to change the attitude and increase r or l via self-conviction. The question is then: how will this affect future behavior?

2.4. **Memory dynamics.** As follows from the assumptions, the instantaneous payoff is a random variable. Its realizations are pairs of real numbers, indicating the reward for the cognitive and impulsive units correspondingly. The distribution is generated by the independent random events: whether the problem is simple or difficult and whether the agent is alert. In addition, the distribution depends on the sign of $I_t - (C_t + h)$, accumulating the history of past success and failures by each unit.

If the sign is negative, so that $I_t < C_t + h$, we say the agent is in the *cognitive regime*: only the cognitive unit is active. Denote the random instantaneous reward in cognitive regime by U, it is described in figure 2.1.

The updated memory is a weighted average of an instantaneous net reward and the old indicator. The relative weight is $0 \leq \beta \leq 1$. So,

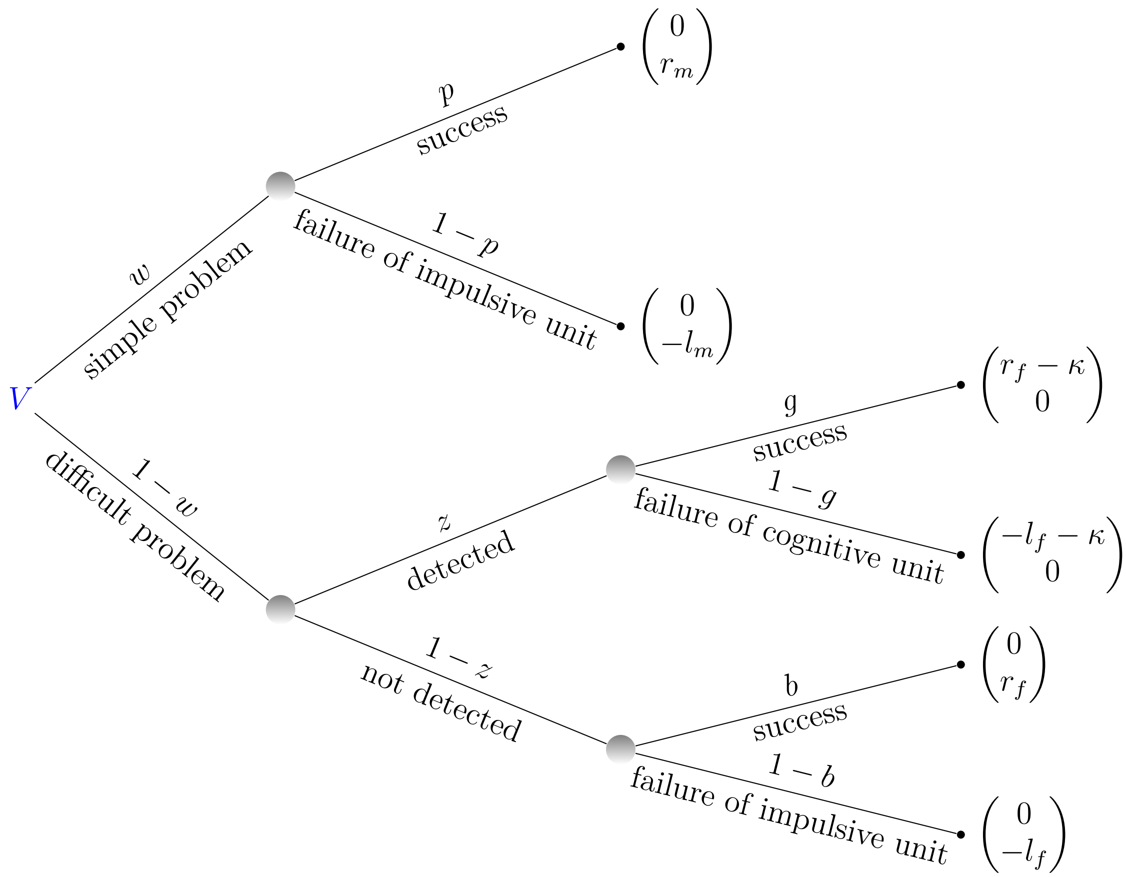

FIGURE 2.2. Random payoff, V, in the impulsive regime. Note that only when the cognitive unit is activated, the payoff is reduced by κ.

when in cognitive regime,

$$\begin{pmatrix} I_{t+1} \\ C_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\beta & 0 \\ 0 & 1-\beta \end{pmatrix} U + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} I_t \\ C_t \end{pmatrix} \tag{1}$$

Now, if $I_t > C_t + h$, the agent is in the *impulsive regime*: impulsive unit is activated unless a difficult problem is detected. Recall, the problem is simple with probability w, otherwise, when the problem is difficult and the agent is alert, so that the difficulty is detected (with probability z), cognitive unit is always activated. The instantaneous reward is the random variable V, described in figure 2.2. The memory is updated according to the same algorithm as in the previous case:

$$\begin{pmatrix} I_{t+1} \\ C_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\beta & 0 \\ 0 & 1-\beta \end{pmatrix} V + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} I_t \\ C_t \end{pmatrix} \tag{2}$$

In case of equality, $I_t = C_t + h$, any convex combination of the two update rules can be used:

$$\begin{pmatrix} I_{t+1} \\ C_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\beta & 0 \\ 0 & 1-\beta \end{pmatrix} (qV + (1-q)U) + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} I_t \\ C_t \end{pmatrix}, \quad q \in [0,1] \tag{3}$$

To sum up, the memory update can be described using the indicator functions, $\mathbb{1}$:

$$\begin{pmatrix} I_{t+1} \\ C_{t+1} \end{pmatrix} = \begin{pmatrix} 1-\beta & 0 \\ 0 & 1-\beta \end{pmatrix} \Big[\mathbb{1}_{I_t = C_t + h}(qV + (1-q)U) + \\ + \mathbb{1}_{I_t > C_t + h} V + \mathbb{1}_{I_t < C_t + h} U \Big] + \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} I_t \\ C_t \end{pmatrix}, \quad q \in [0,1] \tag{4}$$

Next, we want to describe the long-term (asymptotic) behavior of the system: how often the problems are solved correctly, how often each of the modes is being used. And of course, we could then trace the dependence of the long-term behavior on the parameters of the model: probabilities of success of each unit and the payoffs.

Notation.

$$i \stackrel{\text{def}}{=} (r_m, -l_m, r_f, -l_f) \in \mathbb{R}^4$$
$$c \stackrel{\text{def}}{=} i - (\kappa, \kappa, \kappa, \kappa)$$

For the cognitive regime define

$$\pi_{c1} \stackrel{\text{def}}{=} \big(wa, w(1-a), (1-w)g, (1-w)(1-g)\big)$$

For the impulsive regime, let

$$\pi_{i2} \stackrel{\text{def}}{=} \big(wp, w(1-p), (1-w)(1-z)b, (1-w)(1-z)(1-b)\big)$$
$$\pi_{c2} \stackrel{\text{def}}{=} z(1-w)\big(0, 0, g, (1-g)\big)$$

3. Benchmark: amnesiac ($\beta = 0$) long-term behavior

To start, let us analyse the simplest case, with only the short-term memory, i.e., when $\beta = 0$. In addition, in this section we will also set a to unity implying that the cognitive unit always succeeds in solving simple problems. We will (somewhat loosely) call such an agent an amnesiac, referring, of course, only to the very short "emotional" memory associated with past rewards and successes.

In this case, the process has a finite number of states. The state describes which unit was activated (C or I), whether the problem was simple or difficult (M or F) and whether the unit succeeded in its task (S or L). In other words, the set of states can be identified with a set of realizations of payoffs, i.e., possible values of (C_t, I_t). Alternatively, this can be thought of as a set of all the end notes in probability trees V, U.

List the states as follows: CMS, CML, CFS, CFL, IMS, IML, IFS, IFL.

To construct the probability transition matrix, Π, first, observe that according to the model the transition probabilities from each of the states can be of two types, either $(\pi_{c1}, 0, 0, 0, 0)$ or (π_{c2}, π_{i2}).

We illustrate the construction for certain "extreme" values of the inhibition parameter, h, that will give us also (immediately) the long-run behavior of the Markov chain.

If $h > r_f$, then even if the impulsive unit succeeded in solving the difficult problem correctly, the agent switches to a cognitive regime. Assuming, in addition, $r_f > r_m > -l_m > -l_f$, implies that the same is true after any activation of the impulsive regime. If $h > l_f + \kappa > \kappa - r_f$, then even after the cognitive unit makes a mistake, it still prevails. Hence if $h > \max\{r_f, l_f + \kappa\}$ then $(\pi_{c1}, 0, 0, 0, 0)$ is the only row that will appear in the transition matrix in this case. Hence this distribution is the long-run distribution over the eight states: since each row sums up to unity and each column consists of identical elements (as all rows are the same), $\Pi^T = \Pi$ for all T, and so $\lim_{T\to\infty} \Pi^T = \Pi$.

If, on the other hand, $h < \min\{-l_f, -l_m\}$, then each row of the transition matrix is (π_{c2}, π_{i2}), which, similarly, is also the asymptotic distribution in this case. To sum up,

Lemma 1. *If $h > \max\{r_f, l_f + \kappa, r_m, l_m + \kappa\}$, then only cognitive unit is used with positive probability and the distribution over the eight states* CMS, CML, CFS, CFL, IMS, IML, IFS, IFL *is $(\pi_{c1}, 0, 0, 0, 0)$. If $h < \min\{-l_f, -l_m\}$, then the impulsive unit is used unless a difficult problem is detected and the asymptotic distribution is (π_{c2}, π_{i2}).*

Is there an asymptotic distribution in non-extreme cases?

A classical candidate for the asymptotic distribution is the stationary one, i.e., x such that $x\Pi = x$. We will establish existence of such x using Perron-Frobenius theorem, but before the result is established, let us look at an example.

Example 1. Assume (1) $r_f > r_m$; (2) $h > \max\{-l_f, -l_m\}$; (3) $r_m - \kappa + h > 0$; (4) $\min\{r_m, l_m + \kappa, l_f + \kappa\} > h$. Under these assumptions the decision is made by the unit that succeeded in the last round and the control is switched to a different unit after experiencing a loss. Then

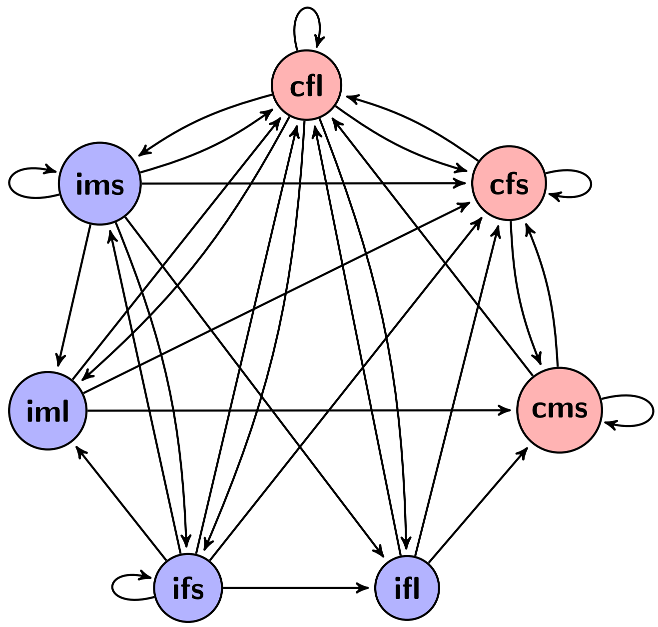

FIGURE 3.1. The Markov chain corresponding to the matrix (5), i.e., assuming $r_m > h > \kappa - r_m$. The nodes represent the states and the arches (edges) correspond to the non-zero transition probabilities. The states colored blue are those where impulsive unit is responsible for solving the problem, and the red states are those where cognitive unit is. The transient state CML is not represented.

the transition matrix, Π is

$$\begin{array}{c|cccc|cccc} & \text{CMS} & \text{CML} & \text{CFS} & \text{CFL} & \text{IMS} & \text{IML} & \text{IFS} & \text{IFL} \\ \text{CMS} & \multicolumn{4}{c|}{\pi_{c1}} & 0 & 0 & 0 & 0 \\ \text{CML} & \multicolumn{4}{c|}{\pi_{c1}} & \multicolumn{4}{c}{0000} \\ \text{CFS} & \multicolumn{4}{c|}{\pi_{c1}} & 0 & 0 & 0 & 0 \\ \text{CFL} & \multicolumn{4}{c|}{\pi_{c2}} & \multicolumn{4}{c}{\pi_{i2}} \\ \text{IMS} & \multicolumn{4}{c|}{\pi_{c2}} & \multicolumn{4}{c}{\pi_{i2}} \\ \text{IML} & \multicolumn{4}{c|}{\pi_{c1}} & 0 & 0 & 0 & 0 \\ \text{IFS} & \multicolumn{4}{c|}{\pi_{c2}} & \multicolumn{4}{c}{\pi_{i2}} \\ \text{IFL} & \multicolumn{4}{c|}{\pi_{c1}} & 0 & 0 & 0 & 0 \end{array} \qquad (5)$$

Π describes a Markov chain, which can be presented as a directed graph, cf. figure 3.1.

It can be verified using figure 3.1 that, excluding CML, there is a path from each state into every other state, so the Markov (sub)chain excluding CML is irreducible. It also implies that from every such state there is a path that ends in the same state and passes through, say, IFS. But then there is also such a loop that has one more transition (edge), as there is a positive probability that the state IFS will be repeated. Hence the chain excluding CML is irreducible and aperiodic, i.e., for T high enough Π^T will have all its columns non-zero apart from CML. Irreducibility and aperiodicity together are sufficient for the Markov chain to have a unique stationary distribution by Perron-Frobenius theorem. This analysis is easy to generalize to include the rest of the cases. We will only assume for the rest of this section that $r_f > r_m$ and $l_f > l_m$, since difficult/novel problems typically generate wider-spread gains. Besides, these assumptions make the arguments shorter. The proof of the following proposition is in the appendix.

Proposition 1. *If $\beta = 0$, then (C_t, I_t) is a finite Markov chain with eight states and it has a stationary (asymptotic) distribution.*

The distribution can be computed explicitly by solving equation $x\Pi = x$ with the relevant transition matrix.[10]

3.1. **Some "comparative statics" for the amnesiac case.** The calculation of the long-run behavior can be simplified, if one re-formulates the problem as a two-state Markov chain with the states being impulsive regime (IR) and cognitive one (CR). Note that the frequency of usage of each regime uniquely determines the frequencies with which each of the units, impulsive and cognitive will be used. These latter frequencies we take as an "observed variable".

Then the transition matrix P for the two regimes is

$$\begin{array}{ccc} & \text{CR} & \text{IR} \\ \text{CR} & \alpha & 1-\alpha \\ \text{IR} & 1-\gamma & \gamma \end{array} \tag{6}$$

where α and γ, both in $[0,1]$, are determined by the relative magnitudes of h and the payoffs. For the example discussed above, probability of staying in the cognitive regime is $\alpha = \pi_{c1} \cdot (1,1,1,0)$ and probability of staying in the impulsive regime is $\gamma = \pi_{i2} \cdot (1,0,1,0)$.

[10]Note though that $\Pi - \mathbf{I}$ (with $\mathbf{I}$ being the identity matrix and Π any $n \times n$ stochastic matrix) is not invertible: unity is an eigenvalue of Π by the Perron-Frobenius theorem and it has the corresponding non-negative eigenvector.

The set of non-trivial non-negative solutions of the system $x(\Pi - \mathbf{I}) = 0$ is typically a line (there is a single "degree of freedom"), and requiring the sum of elements of the solution (x) to be unity uniquely identifies the vector.

If $\alpha = \gamma = 0$, the impulsive regime is followed with probability one by the cognitive regime and vice versa, so the agent spends half of the time in either regime.

The two-state Markov chain is irreducible and is aperiodic unless $\alpha = \gamma = 0$. So if α and γ are not both zero the chain has a stationary distribution $z = (1-q, q)$: $zP = z$, so q solves

$$(1-q)\alpha + q(1-\gamma) = 1-q \tag{7}$$

and thus $q = \frac{1-\alpha}{2-\alpha-\gamma}$.

Hence q is decreasing in α and increasing in γ. Note that α is, for an arbitrary combination of parameters, a product of π_{c1} and a vector of four 0-1 indicators of whether the agent stays in the cognitive regime after the cognitive unit encountered each one of the two problems (simple or difficult) and whether it succeeded in each. Thus, $(1,1,1,0)$ stands for the case where the agent stays in the cognitive regime unless the unit failed to solve a difficult problem. Similarly, γ is the product of π_{i2} and the corresponding four indicators (1 for staying in the impulsive regime and 0 for switching to the cognitive one).

In order to see how α and γ change with h, let us start with the extreme case $h > \min\{r_f, l_f + \kappa\}$. In this case $\alpha = \pi_{c1} \cdot (1,1,1,1) = 1$ and $\gamma = \pi_{i2} \cdot (0,0,0,0) = 0$, which is consistent with agent staying in cognitive regime with probability 1. At the other extreme, $h < -l_f$, $\alpha = \pi_{c1} \cdot (0,0,0,0) = 0$ and $\gamma = \pi_{i2} \cdot (1,1,1,1) = 1$, so $q = 1$: the agent is always in the impulsive mode. Gradually reducing h from $h > \min\{r_f, l_f + \kappa\}$ to $h < -l_f$ increases γ and reduces α: $h < r_f$ implies the vector of indicators in definition of γ has to have 1 as the first entry, $h < r_m < r_f$, implies the first and the third entry has to be unity, further, $h < -l_m < r_m$ implies the first three entries have to be 1, etc. Similarly, reducing h will gradually turn all the indicators in the definition of α to 0 (from 1).

Hence, we have established the following

Lemma 2. *The frequency of being in the impulsive regime weakly decreases with the inhibition parameter h in the amnesiac case, i.e., when $\beta = 0$.*

Similar comparisons can be done by changing the payoffs.

As an example, consider lowering the positive payoff, $r_f (> r_m)$. In this case, if $h > 0$, for r_f low enough, the condition $\kappa - r_f > h > r_f$

will hold. It is easy to check that $\alpha = \gamma = 0$, so the regime is switched every period with probability one.[11]

Interpretation. The individual with low (internal) rewards, r_f, hardly enjoys his own successes. As a result, (if his memory of past payoffs is short-lived) he is always in doubt, perpetually switching from intuition (impulsive regime) to reason (cognitive regime) and back.

Surprisingly, diminishing losses has a different effect, already noted, in part, in lemma 1. If the inhibition is high, $h > 1$, then lowering the loss assures condition $h > \kappa + l_f$, which implies $\alpha = 1$, so $q = 0$ (no matter what value $\gamma \in [0, 1]$ attains). In this case, the individual eventually stays in the cognitive regime. If, in addition, $h > r_f$ (as in lemma 1), then, even if the individual finds himself in the impulsive mode (from the start), he will switch to the cognitive one in the next period. However, if this last inequality is reversed, he might spend several periods (upon success of the impulsive unit) in the impulsive regime, but given $h > l_f + \kappa$ the number of such periods is finite.

If $h < 0$, then diminishing the loss ($|l_f|$) yields $h < -l_f$, so $\gamma = 1$ and thus $q = 1$, for any $\alpha \in [0, 1]$, so the agent is in the impulsive regime with probability one in the long run.

Interpretation. The condition for keeping the amnesiac in the cognitive regime, $h > \min\{r_f, l_f + \kappa\}$, requires inhibition to be high (not surprisingly), but also can be read as making any reward or loss relatively small, perceiving both as relatively unimportant, possibly along the lines of the suggestions given by Kipling (1910) in his poem "If...". The second case, $h < -l_f$, is about the agent with overly powerful impulsiveness, who is driven solely by the "animal spirits".

4. General case, $0 < \beta < 1$, heuristics

The first step is a heuristic one, based on the stochastic approximation techniques developed by Kushner and Yin (1997), also used by Cho and Matsui (2006).

[11] That there is no stationary distribution in the two-state Markov chain IR, CR does not necessarily imply that there is no such distribution for the original eight state chain, note the latter exists by proposition 1.

4.1. **The associated (mean limit) system and its stationary points.** At this point we want to investigate just the behavior of the average of the original process described in equation (4). The conjecture (that will be confirmed in the next section) is that the asymptotic distribution of the memory will be centered around the stationary point of the associated system of ordinary differential equations of sect. 4.1.1.

Notation.

$$\bar{I} \stackrel{\text{def}}{=} \pi_{i2} \cdot i, \quad \bar{C} \stackrel{\text{def}}{=} \pi_{c2} \cdot c, \quad \hat{C} \stackrel{\text{def}}{=} \pi_{c1} \cdot c$$

Observe that $\hat{C}$ equals the single-period expected "payoff" from operating in cognitive regime and $\bar{C} + \bar{I}$ is the average payoff under the impulsive regime.

In what follows we impose assumption 1, which is satisfied if κ, the cognitive cost, is high enough.[12] The assumption assures existence of a non-trivial solution.

Assumption 1. [High cognitive cost] $\hat{C} < \bar{C} - \bar{I}$.

4.1.1. *The associated (mean limit) system.* Define the following system of the ordinary differential equations that describes the behavior of the mean update of C, I:

$$\begin{aligned}
&C'_t = \pi_C \cdot c - C_t, && I'_t = \pi_I \cdot i - I_t, && \text{where} \quad (8)\\
&\pi_C = \pi_{c1}, && \pi_I = (0,0,0,0), && \text{if } I_t < C_t + h\\
&\pi_C = \pi_{c2}, && \pi_I = \pi_{i2}, && \text{if } I_t > C_t + h\\
&\pi_C = (1-q)\pi_{c1} + q\pi_{c2}, && \pi_I = q\pi_{i2}, q \in [0,1], && \text{if } I_t = C_t + h
\end{aligned}$$

Note that once the system reaches the line $I_t = C_t + h$, its behavior becomes a convex combination of the one above the line (with weight q) and that below the line (with weight $1 - q$).

4.1.2. *Stationary solutions.*

Proposition 2. *The stationary solution* (C^*, I^*) *of the following system of the ordinary differential equations* (8) *is*

(K) in case $\bar{I} - \bar{C} < h < -\hat{C}$, $q^* = \frac{\hat{C}+h}{\bar{I}+\hat{C}-\bar{C}}$ *and*

$$C^* = q^*\bar{C} + (1-q^*)\hat{C}, \qquad I^* = q^*\bar{I}$$

(L) in case $h > -\hat{C}$,

$$C^* = \hat{C}, \qquad I^* = 0$$

[12]Interpretation and implications of the assumption are discussed later, in section 6.3.1.

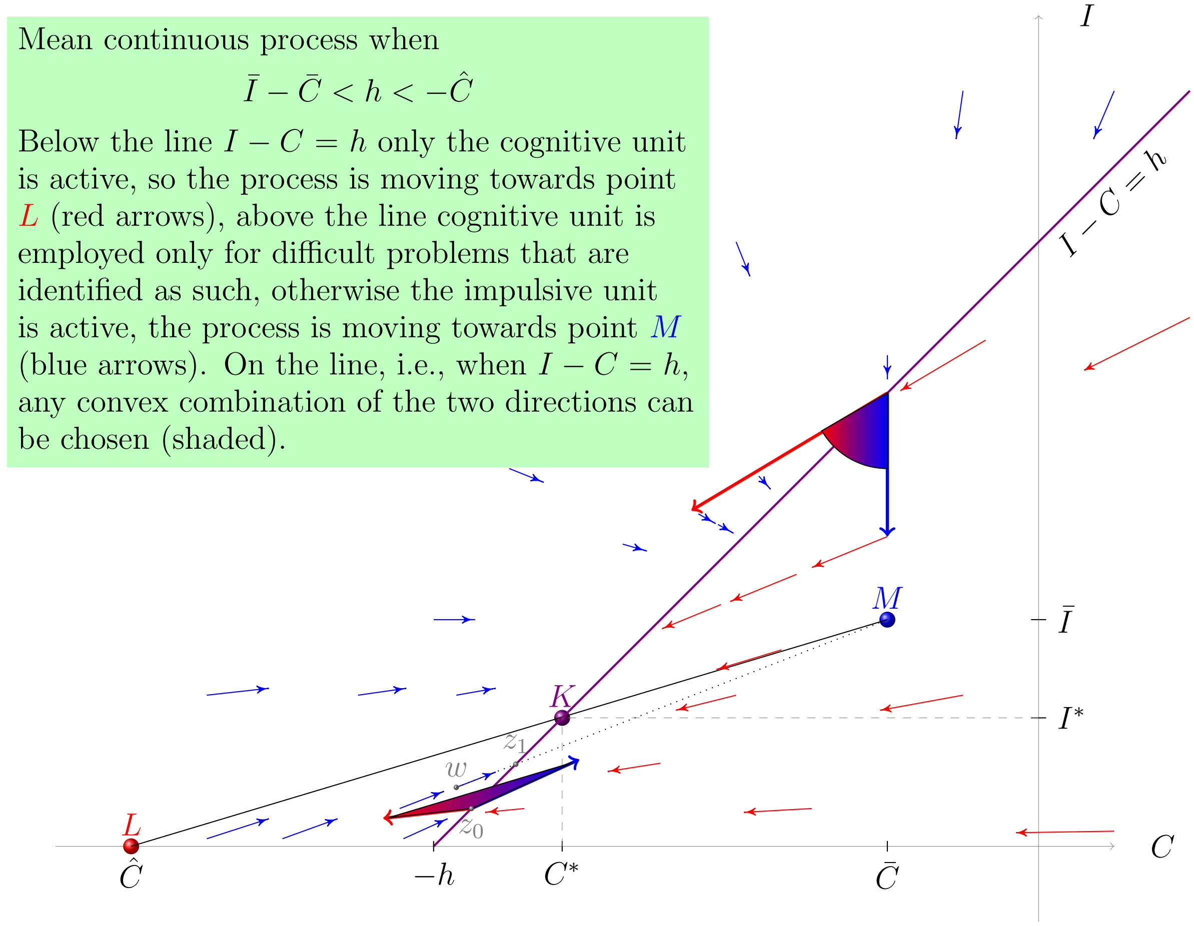

FIGURE 4.1. Stationary point of the mean continuous process is (C^*, I^*).

(M) in case $h < \bar{I} - \bar{C}$,

$$C^* = \bar{C}, \qquad\qquad I^* = \bar{I}$$

Proof. Stationarity implies $C_t = C^* = \pi_C \cdot c$ and $I_t = I^* = \pi_I \cdot i$ for all t. The equality $I^* = h + C^*$, yields then the formulas for C^*, I^* and q in case (K), and $q^* \in [0, 1]$ is assured by the assumption on the range of h and assumption 1.

Similarly, in case (L), the formulas for C^*, I^* satisfy $C^* + h > 0 = I^*$, since then $h > -C^*$. Likewise, in case (M), $C^* + h < I^*$. □

4.1.3. *Stability of the stationary solution.* A direct way to analyze stability of the stationary solution established in proposition 2 is to study the basins of attraction, and for that one can just solve the system of the ODE using lemma 6 in the appendix, which also contains the proof of the stability, prop. 3. Fig. 4.1 illustrates the argument.

Proposition 3. *The stationary solution* (C^*, I^*) *defined in proposition 2 is globally asymptotically stable.*

5. Basic Convergence Result

Having identified the stationary stable points of the average (deterministic) process, let us return to the original problem, i.e., find asymptotic behavior of system (4). Let us re-write the update process

$$\mathbb{1}_{I_t = C_t + h}(qV + (1-q)U) + \mathbb{1}_{I_t > C_t + h}V + \mathbb{1}_{I_t < C_t + h}U$$

in a more convenient way. Introduce Z_t, a random payoff (column) vector in period t, which depends upon the regime. If $I_t < C_t + h$ (cognitive regime), then $Z_t = U$ which is a random variable, whose expected value is

$$\bar{U} = \begin{bmatrix} \pi_{c1} \cdot c \\ 0 \end{bmatrix} = \begin{bmatrix} \hat{C} \\ 0 \end{bmatrix}$$

If $I_t > C_t + h$ (impulsive regime), then $Z_t = V$, whose expected value is

$$\bar{V} = \begin{bmatrix} \pi_{c2} \cdot c \\ \pi_{i2} \cdot i \end{bmatrix} = \begin{bmatrix} \bar{C} \\ \bar{I} \end{bmatrix}$$

Finally, if $I_t = C_t + h$, (both regimes are equally good), $Z_t = qU + (1-q)V$ for some $q \in [0,1]$ (in this case we leave q undefined). By construction, the distribution of Z_t is discontinuous at $I_t = C_t + h$. We want to approximate the asymptotic behavior of the system by the behavior of a smooth system, i.e., by adding a little "noise" in the decision of the control unit. Of course, to make such an approximation meaningful, we want to rely as little as possible on a particular form of noise that we add, so we impose minimal assumptions for a variation of a standard convergence argument to go through. So, first, we introduce the payoff with noise.

For any $\sigma > 0$, conditional on $y_t \overset{\text{def}}{=} C_t - I_t + h$,

$$Z_t^\sigma = \begin{cases} U & \text{with probability } \; \eta^\sigma(y_t) \\ V & \text{with probability } \; 1 - \eta^\sigma(y_t). \end{cases} \tag{9}$$

To make this definition work, we have to assume that η^σ is defined on $\mathbb{R}$ and returns values in the unit interval. Next requirement for approximation to be meaningful is convergence of η^σ to a step function $\mathbb{1}_{y_t > 0}$ as $\sigma \to 0$. Weak convergence of η^σ assures that Z_t^σ converges to Z_t in distribution (as $\sigma \to 0$), but to make arguments simpler, let us assume pointwise (in y) convergence of η^σ to the step function. In addition we will also require η^σ to be continuous and monotone (in its argument y_t) for any $\sigma > 0$.[13]

Next, let X_t denote the two dimensional real vector (C_t, I_t). Then the original system (4) can be written as

$$X_{t+1} = \beta X_t + (1-\beta)Z_t^\sigma + (1-\beta)\zeta_t^\sigma \tag{10}$$

where $\zeta_t^\sigma \stackrel{\text{def}}{=} (Z_t - Z_t^\sigma)$ is the "error of approximation". So, without the last term, the system describes evolution of memory with some noise,[14] i.e., when the control unit makes "mistakes" with probability that decreases with the distance from the threshold, $\{(C_t, I_t)\colon C_t + h = I_t\}$.

Our objective now is to describe the asymptotic distribution of X_t defined by the original system (10). To describe distribution of X_t at any given time t, let us define the underlying probability measure.

Notation. Let $(\Omega, \mathcal{F}, \mu)$ be a probability space such that $X_t \in \mathbb{R}^2$ for non-negative t is defined on it. Let $\mathcal{F}_t^\sigma \subset \mathcal{F}$ be the σ-field generated by $X_t \stackrel{\text{def}}{=} (X_0, Z_0^\sigma, \zeta_0^\sigma, Z_1^\sigma, \zeta_1^\sigma \ldots, Z_{t-1}^\sigma, \zeta_{t-1}^\sigma)$, so that $\{\mathcal{F}_t^\sigma, t \in \mathbb{Z}_+\}$ is an increasing family of σ-fields. Let μ_t be a probability measure defined on $\mathcal{F}_t^\sigma$ for any $t \geq 0$. Let E_t be the expectation with respect to μ_t (or conditional on X_t). Let realizations of all random variables be zero for $t < 0$.

Note that distribution of each of the random variables Z_t^σ depends on $y_t = (1, -1)X_t + h$, hence it is important to understand the asymptotic behavior of y_t. Let us start with its expected value,

$$\mathsf{E}_t(y_{t+1}) = \eta^\sigma(y_t)\hat{C} + (1 - \eta^\sigma(y_t))(\bar{C} - \bar{I}) + h$$

Lemma 3. *For any $\sigma > 0$ define function $F\colon y \mapsto \eta^\sigma(y)\hat{C} + (1 - \eta^\sigma(y))(\bar{C} - \bar{I}) + h$ on $\mathbb{R}$, then its range is $[\hat{C} + h, \bar{C} - \bar{I} + h]$. Let $y^\sigma \in \mathbb{R}$ be a fixed point of F, i.e., it solves*

$$F(y) = y. \tag{11}$$

y^σ exists and is unique.

Proof. Function F maps a real number into a convex combination of $\hat{C}$ and $\bar{C} - \bar{I}$ translated by h. By assumption 1, $\hat{C} < \bar{C} - \bar{I}$. Hence the range. η^σ is a continuous and increasing function, which implies that

[13] Take, for example,

$$\eta^\sigma(y) = \frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{y} e^{\frac{-x^2}{2\sigma^2}}\,dx.$$

[14] Alternatively, one can start with a model including noise, thus taking the system $X_{t+1} = \beta X_t + (1-\beta)Z_t^\sigma$ as the original one.

F is continuous and decreasing. Therefore the difference $F(y) - y$ is strictly decreasing on $\mathbb{R}$, is unbounded at $\pm\infty$, and thus has a single zero, y^σ, which is also the solution of equation (11). □

Fortunately, y^σ also defines the fixed point of the map

$$(C_t, I_t) \mapsto y_t \mapsto \mathsf{E}(Z_t^\sigma | y_t)$$

more precisely,

Corollary 1. *For any $y^\sigma \in \mathbb{R}$ from lemma 3 define*

$$C^\sigma = \eta^\sigma(y^\sigma)\hat{C} + (1 - \eta^\sigma(y^\sigma))\bar{C} \tag{12}$$

$$I^\sigma = (1 - \eta^\sigma(y^\sigma))\bar{I} \tag{13}$$

Then $y^\sigma = C^\sigma - I^\sigma + h$, and

$$z^\sigma \overset{\text{def}}{=} \mathsf{E}(Z_t^\sigma | y_t = y^\sigma) = \begin{bmatrix} C^\sigma \\ I^\sigma \end{bmatrix}$$

Finally, the key to the convergence argument is global stability of y^σ for any given $\sigma > 0$.

Lemma 4. *$y_{t+1} \to y^\sigma$ a.s. as $t \to \infty$ for any $\sigma > 0$.*

Proof. By definition of y_t,

$$\mathsf{E}_t y_{t+1} = \mathsf{E}_t\big((1, -1)[\beta X_t + (1 - \beta)Z_t] + h\big) \tag{14}$$

$$= \beta\big((1, -1)X_t + h\big) + (1 - \beta)\big(\mathsf{E}_t((1, -1)Z_t) + h\big) \tag{15}$$

Note that $(1, -1)X_t + h = y_t$ and $\mathsf{E}_t((1, -1)Z_t) + h = F(y_t)$ with F as defined in lemma 3. It follows that

$$\mathsf{E}_t y_{t+1} = \beta y_t + (1 - \beta)F(y_t) \tag{16}$$

Recall, F is decreasing and by lemma 3 it has a unique fixed point y^σ. It follows that $y_t > y^\sigma \iff F(y_t) < F(y^\sigma) = y^\sigma$, so whenever $y_t > y^\sigma$, it is also true that $y_t > F(y_t)$. Therefore if $y_t > y^\sigma$ then $\mathsf{E}_t y_{t+1} < y_t$ and otherwise $\mathsf{E}_t y_{t+1} \geq y_t$. It follows that the stochastic sequences $\max\{y_t - y^\sigma, 0\}$ and $\max\{y^\sigma - y_t, 0\}$ are both positive supermartingales and hence both converge almost surely (and in distribution) to a limit. Both limits have to be zero, otherwise taking limits in (16) will lead to a contradiction. □

Now we can show that X_t can be approximated by an autoregressive process (AR1), and this is the main result.

Theorem 1. *X_t converges in distribution as $t \to \infty$ to a process with a finite mean and a finite variance:*

$$X_{t+1} = (1 - \beta)\Sigma_{j=0}^{\infty}\beta^j \tilde{Z}_{t-j}^\sigma + T^\sigma$$

where $\tilde{Z}^\sigma$ is distributed as Z_t^σ when $C_t = C^\sigma$ and $I_t = I^\sigma$; and where T^σ converges in distribution to zero as $\sigma \to 0$.

Proof. Recall that by construction Z_t^σ has finite support for any t and is a convex combination of random variables U and V with the corresponding weights $\eta^\sigma(y_t)$ and $(1 - \eta^\sigma(y_t))$. By lemma 4 and continuity of η, distribution of Z_t^σ converges (a.s.), as $\eta^\sigma(y_t) \to \eta^\sigma(y^\sigma)$. Denote the limiting random variable by $\tilde{Z}^\sigma$, it follows that its expectation at any time is z^σ as defined in cor. 1, it is finite and so, obviously is its variance.

The noise term, ζ_t^σ is also a random variable with a finite number of possible realizations and is bounded by the highest (in absolute value) realization of U and V. Hence $T^\sigma = (1-\beta)\Sigma_{j=0}^{\infty}\beta^j\zeta_{t-j}^\sigma$ is bounded and by construction converges to zero in distribution when $\sigma \to 0$. □

Remark 1. It immediately follows that the expectation of X_t is finite, as the range of possible values is bounded. Given the realizations of U, V and hence $\tilde{Z}^\sigma$ and its distribution, which is determined by $\eta^\sigma(y^\sigma)$, one can compute the approximate (up to the error term T^σ) asymptotic distribution of X_t directly from the statement of the proposition.

As mentioned above (by construction of η^σ), when $\sigma \to 0$, the error term converges to zero (in distribution). Now it is left to find out what happens to the distribution of $\tilde{Z}^\sigma$ as $\sigma \to 0$, or at least, to its mean. Apparently it converges to the stationary point of the associated mean ODE that was derived in prop. 2, at the stage of the "heuristic" analysis. The proof is in the appendix.

Lemma 5. *$z^\sigma = (C^\sigma, I^\sigma) \to (C^*, I^*)$ as $\sigma \to 0$.*

Now, combining lemma 5 with theorem 1, we get an explicit description of the asymptotic distribution.

Corollary 2. *If $t \to \infty$, $\sigma \to 0$ and $\beta \to 1$, then the distribution of X_t is normal with mean (C^*, I^*), defined in proposition 2.*

Remark 2. The order of limits matters: if $\sigma \to 0$ is taken first and $t \to \infty$ thereafter, the result might differ. Indeed, if $\sigma \to 0$ the fixed point of (11), has to converge to zero, but for any $\sigma > 0$, $\eta^\sigma(0)$ depends on the construction of η, using the example from footnote 13, $\eta^\sigma(0) = \frac{1}{2}$, so the stationary point in the limit has to be both half-way between $\bar{U}$ and $\bar{V}$ and satisfy $y = 0$, which is possible only for a knife-edge choice of parameters: $\frac{1}{2}(\hat{C} + \bar{C} - \bar{I}) + h = 0$. However, the last two limits can be taken in any order.

Remark 3. Under assumptions of corollary 2 it is possible to characterize the asymptotic behavior even more precisely, provided $(\hat{C}, 0) \neq (\bar{C}, \bar{I})$. For example, in case (L) of prop. 2, i.e., if $h > -\hat{C}$ the corollary implies that the average of the asymptotic memory is $(\hat{C}, 0)$, hence the individual has to spend most of the time in the cognitive regime. Indeed, otherwise, if a positive fraction of time were to be spent in the impulsive mode, the average would have been shifted towards $(\bar{C}, \bar{I})$. Similarly, in case (M) of prop. 2 the agent spends (asymptotically) most of the time in the impulsive regime. Finally, in the complementary case (K), he has to spend a positive fraction of time in either regime, corresponding to q^* described in prop. 2.

6. Empirical implications and discussion

In this section we assume that the assumptions of corollary 2 are satisfied, so, in particular, β is close to unity, i.e., the memory (or recall) is almost perfect.

6.1. **Inhibition of impulses and successful thinking.** One of the most immediate consequences of the main result and its corollary is that if $h > 0$ and $\hat{C} > 0$ (so that $h > -\hat{C}$), then the average of the asymptotic distribution will be $(\hat{C}, 0)$, i.e., that of operating the cognitive regime only.[15] So, some self-control ($h > 0$) and a rather successful cognitive function with positive expected reward (net of thinking cost) implies cognitive unit is the only one operating in the long run. Is this a reasonable implication? One could think of a diligent accountant as an example: all problems, no matter how mundane are given undivided attention, no intuitive guesses are made, all the work is done carefully and thoughtfully. This definitely requires a certain amount of self-control and a relatively successful operation. The accountant might not even be able to recall what it is like to rely on intuition in his work! No "silly" mistakes are made, all the decisions are fully rational.

This does not imply however that the expected payoff from thinking is the highest, intuition is shut down just because cognitive unit performs *sufficiently well* and is being favored by the executive unit (due to inhibition).

6.2. **Comparative statics: changes in nature and nurture.** Now we want to find out how the asymptotic (predicted) average, (C^*, I^*), is affected by the parameters. The proof is in the appendix.

[15]Returning to figure 4.1 such a scenario is possible when the segment LM is shifted to the right into the positive orthant. Cf. also remark 3.

Proposition 4. *The asymptotic frequency of using cognitive regime increases if any one of the conditions below holds*

(A) *inhibition parameter, h, increases;*
(B) *cost of thinking, κ, decreases;*
(C) $z > \frac{b}{b+g}$ *and r_f increases;*
(D) $z > \frac{1-b}{2-b-g}$ *and l_f falls.*

Surprisingly, some of the results from the benchmark case carry over to this case, at least in terms of averages.

Observe that if $h > l_f + \kappa$, as is required in the amnesiac case for cognitive regime to prevail, and assuming $l_f > l_m > 0$, the condition $h > \hat{h} \stackrel{\text{def}}{=} -\hat{C}$ is satisfied (because $l_f + \kappa > -\hat{C}$). Hence lowering losses in this case has the same effect (on the average) as in the benchmark.

In contrast to the benchmark, however, increasing rewards can increase the range of h for which cognitive regime prevails, as follows from proposition 4.

Interpretation. Case (A). Inhibition, or strength of the executive control, h, is given by "nature" and, as one would expect, the model predicts that the higher the inhibition, the more the cognitive regime is used and hence fewer decisions are made impulsively.

Case (B) demonstrates that increasing the cost of thinking decreases the frequency of its use. Importantly, no "conscious" decision is involved and no meta-calculations are performed to assess the cost-benefit analysis of thinking. The result is caused by the reinforcement-learning-based trigger activating the best-performing unit.

Cases (C)-(D). Even if inhibition is not very high, but the agent is alert and identifies difficult problems frequently, then increasing the "carrot" and reducing the "stick" enhances more thoughtful or contemplative decisions, which are made by the cognitive unit.

6.3. **Discussion.**

6.3.1. *The high cognitive costs assumption.* Recall that all the results we got are subject to assumption 1: $\bar{I} < \bar{C} - \hat{C}$.

First, to justify the name of the assumption, notice that the left hand side of the inequality is independent of κ, whereas the difference on the right hand side is increasing in κ:

$$\begin{aligned}\bar{C} - \hat{C} = (\pi_{c2} - \pi_{c1})c &= wa(\kappa - r_m) + w(1-a)(\kappa + l_m) \\ &+ (1-w)g(1-z)(\kappa - r_f) + (1-w)(1-g)(1-z)(\kappa + l_f)\end{aligned}$$

While there are many reasons to believe that the cost κ for thinking about a problem is not negligible (see section 1.1.1 in the introduction), thus, in a way, supporting the assumption, one might still wonder what happens if it is violated.

By inspecting the proof of proposition 2, one can easily see that the same argument holds in the two extreme cases. If (M) is true, but not (L) (the stationary point then is $(\bar{C}, \bar{I})$, that of the impulsive regime); and if (L) is true, but not (M) (the stationary point then is $(\hat{C}, 0)$, that of the cognitive regime). However, if both conditions (L) and (M) hold, then there are two stable points with the region of attraction being the corresponding half-plane (above and below the line $C + h = I$ for $(\bar{C}, \bar{I})$ and $(\hat{C}, 0)$ correspondingly). Most likely, in the latter case the results will have to be formulated for the memory starting in some "locality" of the corresponding stable point, although, apart from this case, we conjecture, the results would remain (qualitatively) the same.

6.3.2. *Rethinking the carrot-and-stick policy.* The reader can probably confirm that many academics tend to raise their children in a "supportive" environment, with inflated praise for success and very mild punishments. The grown offsprings rather often decide to become academics themselves. If such an outcome is, at least partially, due to the supportive environment, this should be puzzling at first, at least in the view of the classical mechanism design literature. There payoffs can be "normalized" so that only the gap between the best and the worst payoff can affect behavior, their shift being of no consequence. However this observation is in accord with our findings: cases (C)-(D), proposition 4.

As an implication, if one is to motivate an agent to make more thoughtful decisions and thus avoid "silly" mistakes coming from operating on "auto-pilot", increasing the carrot and decreasing the stick on a regular basis might do the trick (provided our assumptions hold). And this is despite the fact that "the principal" is incapable of rewarding cognitive and impulsive units directly (the payoff depends only on success or failure of solving a problem).

6.3.3. *An uptight dimwit and a laid-back wise-guy.* The model is general enough to capture a variety of scenarios. In particular, recall that no assumptions have been made about cognitive unit being more successful in solving problems of any kind.

To illustrate, graph 4.1 depicts a story of an uptight dimwitted DM: in intuitive regime he is more successful on average. Indeed, when only cognitive regime is used, his average reward is $\hat{C} = \pi_{c1} c$, corresponding to the red dot in the graph. However, in the impulsive regime it is $\bar{I} + \bar{C} = \pi_{i2}i + \pi_{c2}c$, the sum of the coordinates of the blue dot. But $\bar{C} + \bar{I} > \bar{C} - \bar{I}$ since in this case $\bar{I} > 0$ and $\bar{C} - \bar{I} > \hat{C}$ by assumption 1. So, one can call the DM "dimwitted". Note that despite the relative success of the impulsive regime, the DM is "stuck" in a stable point K that describes his asymptotic average and this is because his inhibition h is high, which forces him to think too much! Being less uptight (lowering h if possible) and thus "overthinking" less, can help the DM to gain higher average reward, and if this is a measure of well-being, then it should make him happier too.

It is not hard to construct an example of a DM with high thinking capabilities and poor intuition, call him a laid-back wise-guy. Take, for example a DM with $\bar{I} < 0$ to reflect the poor intuition, this holds if the impulsive unit is rarely successful in solving either type of problem: $p < \frac{l_m}{l_m + r_m}, b < \frac{l_f}{l_f + r_f}$. In addition, assume that his cognitive unit is finding the right solution sufficiently often, $a > \frac{l_m}{l_m + r_m}, g > \frac{l_f}{l_f + r_f}$, thus assuring $\hat{C} > \bar{C}$. Keep the cognitive costs high to assure that assumption 1 is satisfied ($\hat{C} - \bar{C} > -\bar{I}$). Finally, the inhibition parameter h has to be in the intermediate region, $\bar{h} < h < \hat{h}$, so the wise-guy is rather laid-back. Since assumption 1 still holds, our results apply, hence with perfect recall the asymptotic distribution will be centered around the intersection point between the line $C + h = I$ and the segment connecting $(\hat{C}, 0)$ and $(\bar{C}, \bar{I})$. In this case, however, the average payoff in cognitive regime, $\hat{C}$ is higher than that in the impulsive, $\bar{C} + \bar{I}$ (by construction, $\hat{C} > \bar{C} > \bar{C} + \bar{I}$), thus the relaxed wise-guy could have had a higher average reward with a higher inhibition h (by proposition 4), thinking more often and so erring less.

7. Conclusions

We have offered a description of a technology of decision making involving a simple mechanism with three units: cognitive (contemplative), impulsive (fast, intuitive) and executive control triggering either of the two units to make the decision.

In a stationary Markovian environment, the model yields predictions of individual behaviour summarised in proposition 4: higher inhibition, lower cost of thinking and, in case of high alertness, higher rewards for solving novel problems as well as lower losses, all enhancing more frequent thinking, provided the memory is almost perfect. This might be a welcome change, at least for the laid-back wise-guy described just above, since "...the sleep of reason produces monsters",[16] and so less unwanted purchases are made and less empty promises are being trusted. However, in some cases, as in the story of the uptight dimwit, an awakened mind might be a worse alternative.

We also show that normalizing payoffs by shifting them by a constant is not as innocuous as one might assume and can have an affect on the resulting behavior.

Further, we would like to understand better the choice of action and this, among others, should include the process of "acquiring experience", which involves remembering the successful choices made by the cognitive unit in the past and picking them without thinking in the future, conditional on a particular trigger that encodes the state sufficiently well. This is the avenue for future research.

Appendix A. Proofs

Proof of prop. 1. (C_t, I_t) is a finite Markov chain with eight states (listed in lemma 1) by construction.

The existence and uniqueness of the stationary distribution in the extreme cases is by lemma 1. It remains to consider the complementary case: $-l_f < h < \max\{r_f, l_f + \kappa\}$.

First, note that $h > -l_f$ implies that a failure in the impulsive regime (after solving at least a difficult problem) will cause the control unit to switch to the cognitive regime, i.e., the transition probability in state IFL is $(\pi_{c1}, 0, 0, 0, 0)$. So once the agent is in the impulsive mode, there is a positive probability to switch to the cognitive regime. Thus, in the graph of the chain, there is a path from any impulsive state IMS, IML, IFS, IFL to any cognitive state CMS, CFS, CFL (excluding CML).

Assume now that $\max\{r_f, l_f + \kappa\} = r_f$ and $r_f > h > l_f + \kappa$. $h > l_f + \kappa$ implies that the agent will remain in the cognitive regime even after making a mistake, so once in the cognitive regime, he will remain there with probability 1, hence there is no path (in the graph of the chain) from a cognitive state to an impulsive state. Thus in this case the subchain consisting of the three "cognitive states" CMS, CFS, CFL is recurrent and aperiodic (the latter is because there is a positive probability that any of the three states will be repeated).

Alternatively, $h < l_f + \kappa$, so a failure of the cognitive unit to solve a difficult problem triggers a switch to the impulsive mode, i.e., the CFL row of the transition matrix is (π_{c2}, π_{i2}). Thus, in the graph of the chain there is a path from CFL (and so also from CMS, CFS) to any of the impulsive states IML, IMS, IFL, IFS. It follows that the seven states

[16] *El sueño de la razon produce monstruos*, as claimed by Francisco de Goya.

comprise an irreducible Markov chain. It remains to show aperiodicity for this case.

If $\kappa - r_f > h > r_f$ then the agent switches from one regime to the other with probability one every period. Thus half of the time will be spent in each regime, so there is a positive probability that a state, say CFL, will be repeated.

If $h < r_f$ then a success in the impulsive regime will keep the agent in that regime for the next round, i.e., the transition probability distribution in the state IFS is (π_{c2}, π_{i2}). Thus, in particular, there is a positive probability that the state IFS will be reached twice consecutively.

It follows that in this case the seven states (all but CML) comprise an irreducible aperiodic chain.

By the Perron-Frobenius theorem an irreducible aperiodic chain has a stationary distribution. CML can not be reached from any other state (it is transient), hence it is not in the support of that distribution (in any of the cases). □

Lemma 6. *Fix a locally integrable real-valued bounded function A defined on $\mathbb{R}$ and $b > 0$. The differential equation $f'_t = A_t - bf_t$ has the following solution*

$$f_t = e^{-b(t-t_0)} f_{t_0} + \int_{t_0}^{t} e^{-b(t-s)} A_s ds \tag{17}$$

Proof. By direct computation. □

Proof of prop. 3. Using lemma 6 with $b = 1$ and A_s as $\pi_C \cdot c$ and $\pi_I \cdot i$ correspondingly, the solution in each of the cases should be of the form:

$$C_t = e^{-(t-t_0)} C_{t_0} + \pi_C \cdot c(1 - e^{-(t-t_0)}) \tag{18}$$

$$I_t = e^{-(t-t_0)} I_{t_0} + \pi_I \cdot i(1 - e^{-(t-t_0)}) \tag{19}$$

Assume now $h > -\hat{C}$, as in case (L) of proposition 2. If $C_{t_0} + h > I_{t_0}$, then equations (18)-(19) imply that there is some $\hat{t} \leq \infty$ such that starting from t_0 up to $\hat{t}$

$$C_t = e^{-(t-t_0)} C_{t_0} + \hat{C}(1 - e^{-(t-t_0)}) \tag{20}$$

$$I_t = e^{-(t-t_0)} I_{t_0} \tag{21}$$

For this system the derivative of the difference, $C_t - I_t$, equals

$$e^{-(t-t_0)}(\hat{C} - (C_{t_0} - I_{t_0})) \tag{22}$$

which is non-negative, whenever $\hat{C} \geq C_{t_0} - I_{t_0}$. This implies that if $\hat{C} > C_{t_0} - I_{t_0} > -h$, then $C_t + h > I_t$ for all $t \geq t_0$, hence (20)-(21)

applies for all $t \geq t_0$ (so $\hat{t} = \infty$), and the asymptotic solution is

$$\lim_{t\to\infty} e^{-(t-t_0)} I_{t_0} = 0 \tag{23}$$
$$\lim_{t\to\infty} e^{-(t-t_0)} C_{t_0} + \hat{C}(1 - e^{-(t-t_0)}) = \hat{C}$$

In other words, the basin of attraction of the above solution includes all pairs of starting values C_{t_0}, I_{t_0} that satisfy $\hat{C} > C_{t_0} - I_{t_0} > -h$. The set of such values is non-empty as $h > -\hat{C}$ in this case.

Consider now the starting value for which the derivative of the difference (22) is negative, i.e., $C_{t_0} - I_{t_0} > \hat{C}$. In this case, the difference $C_t - I_t$ will fall with time, t, until it reaches (at some finite time $\tilde{t}$) the value $\hat{C}$, and from thereon the difference will stay constant, since its derivative (22) is zero, while the solution will converge according to (23) to the stationary one in this case (as in proposition 2.(L)).

Finally, if $I_{t_0} - C_{t_0} > h > -\hat{C}$, then (18)-(19) imply that there is $\hat{t} \leq \infty$ such that starting from t_0 up to $\hat{t}$

$$C_t = e^{-(t-t_0)} C_{t_0} + \bar{C}(1 - e^{-(t-t_0)}) \tag{24}$$
$$I_t = e^{-(t-t_0)} I_{t_0} + \bar{I}(1 - e^{-(t-t_0)}) \tag{25}$$

Since the derivative of the difference $I_t - C_t$ for $0 < t < \hat{t}$ equals

$$e^{-(t-t_0)}(\bar{I} - \bar{C} - (I_{t_0} - C_{t_0})) \tag{26}$$

the difference should be decreasing, i.e., $\bar{I} - \bar{C} < I_{t_0} - C_{t_0}$, because $I_{t_0} - C_{t_0} > h > -\hat{C}$, and by assumption 1, $-\hat{C} > \bar{I} - \bar{C}$. Hence from a point $I_{t_0} - C_{t_0} > h$ the difference will decrease, and there is a finite $\hat{t}$ such that $I_{\hat{t}} - C_{\hat{t}} = h$. At that point the left derivative, given by (26) is strictly negative, and the right derivative, which is a convex combination of (26) and the negative of (22), is strictly negative as well. It follows that $I_t - C_t$ will fall below h after $\hat{t}$ and hence follow the path analyzed above (case $C_{t_0} + h > I_{t_0}$).

To sum up, if $h > -\hat{C}$, then the solution from proposition 2, case (L) is globally asymptotically stable.

Assume now $\bar{I} - \bar{C} > h$, as in case (M) of proposition 2.

If $\bar{I} - \bar{C} > I_{t_0} - C_{t_0} > h$, then the path evolves according to (24)-(25) so that the difference, $I_t - C_t$, will be increasing in t (since its derivative (26) is positive) and hence it will be above h for all $t > t_0$. If the difference starts above $\bar{I} - \bar{C}$, then it initially falls down to this level and then remains non-decreasing, hence, also stays above h for all $t > t_0$. If the difference starts below h, the path C_t, I_t evolves according to (20)-(21) and the difference $I_t - C_t$ increases up to h (its derivative by (22) is $e^{-(t-t_0)}(C_{t_0} - I_{t_0} - \hat{C})$, which is positive, since

$C_{t_0} - I_{t_0} > -h > \bar{C} - \bar{I} > \hat{C}$, where the last inequality follows from assumption 1). Again, at h, the sign of the derivative is unchanged, hence the path crosses the line and follows (24)-(25), so, the difference stays above h, as was shown just before. Thus, independently of the initial point, the solution converges to

$$\lim_{t\to\infty} e^{-(t-t_0)}C_{t_0} + \bar{C}(1 - e^{-(t-t_0)}) = \bar{C} \tag{27}$$

$$\lim_{t\to\infty} e^{-(t-t_0)}I_{t_0} + \bar{I}(1 - e^{-(t-t_0)}) = \bar{I} \tag{28}$$

are all pairs C_{t_0}, I_{t_0} satisfying $I_{t_0} - C_{t_0} > h$, provided $\bar{I} - \bar{C} > h$. Therefore the stationary solution described in case (M) of proposition 2 is globally asymptotically stable.

Finally, assume $\bar{I} - \bar{C} < h < -\hat{C}$ as in case (K) of prop. 2.

If the starting point is such that $C_{t_0} + h > I_{t_0}$, then the derivative (22) of the difference, $C_t - I_t$, is negative, therefore, at some finite point $\hat{t}$ in time $C_{\hat{t}} - I_{\hat{t}}$ will equal $-h$. Similarly, if $I_{t_0} - C_{t_0} > h$ by (26), this difference will, at some finite time $\hat{t}$ reach h, so that $I_{\hat{t}} - C_{\hat{t}} = h$. Therefore, it is sufficient to consider the starting point $t_0 = \hat{t}$, i.e., $C_{t_0} + h = I_{t_0}$. Then for some q

$$\lim_{t\searrow t_0} I'_t - C'_t = -I_{t_0} + q\bar{I} + C_{t_0} - q\bar{C} - (1-q)\hat{C} \tag{29}$$

$$= q\bar{I} - q\bar{C} - (1-q)\hat{C} - h \tag{30}$$

If $q = q^*$ from proposition 2, then the derivative (29) is zero and the solution converges to the stationary point (C^*, I^*) specified in proposition 2 for this case (K).

If $q \neq q^*$, then, we claim, there is a subsequence $z_n = (C_{t_n}, I_{t_n})$, on the line, $C_{t_n} + h = I_{t_n}$, that converges to the same point. Here is why. Consider a point on the line, denote it by $z_0 = (C_{t_0}, I_{t_0}) \neq (C^*, I^*)$ (cf. fig. 4.1). As follows from the definition of the mean ODE, eq. (8), for any $q \in [0, 1]$, the right derivative at z_0 points to a direction which is a convex combination of the two systems: first, for the half-plane above the line, (24)-(25), and second, below the line, (20)-(21). As it follows from these equations, the first direction is towards point $(\bar{C}, \bar{I})$, denoted by M on the graph, in fig. 4.1, and the second direction is towards point $(\hat{C}, 0)$, L on the graph. Hence the path moves into the simplex formed by z_0, L, M. Since $q \neq q^*$, the path will not follow the line towards K, but, rather, move towards a different point on the segment containing L and M. Assume it moves towards a point above the line $I = C - h$ (the argument for the complementary case is similar), and so passes through a point, say, w (illustrated in fig. 4.1 as well). But at any point above the line the process is governed by the system (24)-(25) and so moves towards point M, till it reaches the line $I = C - h$. The point, z_1, where it hits the line therefore is on the intersection of the segment starting at w and ending at M (this is the direction of the process (24)-(25)) and the line, precisely, the segment z_0K. Since w is in the simplex formed by z_0, L, M, so is the segment connecting w with the simplex corner, M. K, the candidate stationary point, too, is in the simplex being a convex combination of L and M by case (K) of proposition 2, so the segment z_0K belongs to the simplex. Hence, the point of intersection z_1 between the segments z_0K and wM belongs to the simplex as well. This point, z_1, is a convex combination of z_0 and K and does not coincide with z_0, since $z_0 \neq K$, therefore z_1 is closer to K than z_0.

Let now $(C_{t_1}, I_{t_1}) = z_1$ and find the next point where the process hits the line $I = C - h$ by repeating the same argument.

It follows that the subsequence so constructed converges to K.

If the process starts at K, then it remains there provided $q = q^*$. For a different q, its direction is either towards M or towards L, i.e., it stays on the LM segment. But from any point on the segment it is either above the line, hence follows $(24) - (25)$ pointing to M, or below the line, hence follows $(20) - (21)$ pointing to L, thus returning it to K in both cases. □

Proof of lemma 5. Consider (C^*, I^*) from proposition 2. Let $y^* = C^* - I^* + h$. As a first step, we show that $y^\sigma = C^\sigma - I^\sigma + h$ converges to y^* as $\sigma \to 0$.

For any $\sigma > 0$, by cor. 1, y^σ satisfies

$$\eta^\sigma(y^\sigma)\hat{C} + (1 - \eta^\sigma(y^\sigma))(\bar{C} - \bar{I}) + h = y^\sigma \tag{31}$$

Define now F as a function of two variables:

$$F\colon (y, \sigma) \mapsto \eta^\sigma(y)\hat{C} + (1 - \eta^\sigma(y))(\bar{C} - \bar{I}) + h$$

It is decreasing in y and, by the pointwise convergence of η^σ, for $y > 0$ it is increasing in σ, while for $y < 0$ it decreases in σ.

Pick some $\sigma_0 > 0$. Then, by lemma 3, there is a unique y^{σ_0} such that $F(y^{\sigma_0}, \sigma_0) = y^\sigma$.

Assume $y^{\sigma_0} > 0$. For any $0 < \sigma_1 < \sigma_0$, $F(y^{\sigma_0}, \sigma_1) < F(y^{\sigma_0}, \sigma_0) = y^{\sigma_0}$, hence $F(y^{\sigma_0}, \sigma_1) - y^{\sigma_0} < 0$. The right hand side of the inequality is a decreasing function of y, hence its zero, y^{σ_1}, should be below y^{σ_0}.

So, whenever $y^\sigma > 0$, it is increasing in σ.

If $y^{\sigma_0} < 0$, then, similarly, for any $\sigma_1 > \sigma_0$, $F(y^{\sigma_0}, \sigma_1) < F(y^{\sigma_0}, \sigma_0) = y^{\sigma_0}$, hence $F(y^{\sigma_0}, \sigma_1) - y^{\sigma_0} < 0$, and so, again, $y^{\sigma_1} < y^{\sigma_0}$. Hence whenever $y^\sigma < 0$, it is decreasing in σ.

By lemma 3, y^σ belongs to a compact $[\hat{C}+h, \bar{C}+\bar{I}+h]$, so any infinite sequence $(y^\sigma)_{\sigma\to 0}$ has a converging subsequence.

The limiting point depends on the parameters.

Assume $\hat{C}+h>0$, as in case (L) of proposition 2, so that $y^*=\hat{C}+h>0$. But then by lemma 3, $y^\sigma > y^* > 0$ for any $\sigma>0$, and by the argument above, $F(y^\sigma,\sigma)=y^\sigma>0$ is decreasing as $\sigma\to 0$. In addition, by pointwise convergence of η^σ, as $\sigma\to 0$, for any $y>0$, $F(y,\sigma)\searrow \hat{C}+h=y^*$, so $y^\sigma\searrow y^*$.

Similarly, if $\bar{C}-\bar{I}+h<0$, as in case (M) of proposition 2, so that $y^*=\bar{C}-\bar{I}+h<0$, the corresponding sequence y^σ increases and in the limit approaches y^*.

So, the remaining case is (K), i.e., when $y^*=0$. Then we claim any converging subsequence of $(y^\sigma)_{\sigma\to 0}$ has $y^*=0$ as its limit. Assume to the contrary that one can choose a subsequence that converges to some $x\neq 0$. But the pointwise convergence of η^σ to the step function implies that $F(x,\sigma)$ approaches $\hat{C}+h<0$ if $x>0$ and so for σ small enough $F(x,\sigma)<0<x$ and, similarly, for $x<0$ if σ is small enough $F(x,\sigma)>0>x$, hence $x\neq 0$ can not be a limit of the sequence satisfying $y^\sigma=F(y^\sigma,\sigma)$ for all $\sigma>0$.

The rest of the proof follows by corollary 1. □

Proof of prop. 4. As follows from proposition 2, (C^*,I^*) belongs to the interval connecting $(\hat{C},0)$ and $(\bar{C},\bar{I})$. The closer it is to the first point, the more often the cognitive regime (and hence the cognitive unit) is used. So, it is sufficient to show that the point (C^*,I^*) moves closer to $(\hat{C},0)$ under each of the conditions stated.

By proposition 2, there are three cases to consider.

First, if $h>\hat{h}\stackrel{\text{def}}{=}-\hat{C}$, then $(C^*,I^*)=(\hat{C},0)$. So, in this case, one has to assure that the inequality is not violated, as a result of a change, i.e., either h grows or the threshold, $\bar{h}$ falls.

Second, if $h<\bar{h}\stackrel{\text{def}}{=}\pi_{i2}i-\pi_{c2}c=\bar{I}-\bar{C}$, then $(C^*,I^*)=(\bar{C},\bar{I})$, so in this case (C^*,I^*) is the furthest it can be from the other end of the segment, $(\hat{C},0)$. Moving h into a different region (violating the inequality) can move (C^*,I^*) closer to $(\hat{C},0)$ according to proposition 2. This can be accomplished by either increasing h or by lowering the threshold $\bar{h}$.

Finally, consider the intermediate case, $\bar{h}<h<\hat{h}$. The relevant range is non-empty by assumption 1. Then $q=\frac{h+\hat{C}}{\bar{I}-\bar{C}+\hat{C}}=\frac{h-\hat{h}}{\bar{h}-\hat{h}}$ from proposition 2, defines the average of the asymptotic distribution of the indicators: $(C^*,I^*)=q(\bar{C},\bar{I})+(1-q)(\hat{C},0)$.

Lemma 7. *Assume* $\bar{h} < h < \hat{h}$*. Then* $\frac{h-\hat{h}}{\bar{h}-\hat{h}}$ *is decreasing in* h *and is increasing in* $\bar{h}$ *and* $\hat{h}$*.*

Proof. The first claim holds by assumption 1: $\bar{h}-\hat{h} < 0$; the second one is true since $h - \hat{h} < 0$; and the third one holds because $h - \bar{h} > 0$. □

It follows that any factor decreasing $\hat{h}$ and $\bar{h}$ or an increase in h can not move (C^*, I^*) away from $(\hat{C}, 0)$. So, claim (A) follows.

To demonstrate the rest, we need two auxiliary statements.

Lemma 8. $\hat{h} \overset{\text{def}}{=} -\hat{C}$ *decreases in rewards* r_f *and* r_m *and increases in the cost,* κ*, and in the losses,* l_f*,* l_m*.*

Proof. The assertion directly follows from the definition:

$$-\hat{C} = w\big(a(\kappa - r_m) + (1-a)(\kappa + l_m)\big) + (1-w)\big(g(\kappa - r_f) + (1-g)(l_f + \kappa)\big)$$

□

Lemma 9. $\bar{h} \overset{\text{def}}{=} \bar{I} - \bar{C}$ *increases with* r_m, κ *and falls with* l_m*. It increases with* r_f *iff* $(1-z)b > zg$ *and falls with* l_f *iff* $(1-g)z < (1-z)(1-b)$*.*

Proof. The assertions follow directly from the computation below:

$$\begin{aligned}\bar{I} - \bar{C} =& wpr_m - w(1-p)l_m + (1-w)(1-z)br_f - (1-w)(1-z)(1-b)l_f \\ &+ (1-w)z((1-g)l_f + g(\kappa - r_f))\end{aligned}$$

□

So, statement (B) follows directly from the two lemmata above, since a decrease in κ decreases both $\hat{h}$ and $\bar{h}$.

(C): Since $\hat{h}$ decreases in r_f by lemma 8, it is sufficient to assure that $\bar{h}$ falls with r_f, which is true if $(1-z)b < zg$ by lemma 9, thus the statement.

(D): Similarly, by lemma 8, $\hat{h}$ increases in l_f, and by lemma 9, $\bar{h}$ increases in l_f if $(1-g)z > (1-z)(1-b)$. □

References

Alos Ferrer, C. (2013). Think, but not too much: A dual-process model of willpower and self-control. Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Behavioral Economics, Underlying Principles, No. D05-V1.

Bari, A. and T. W. Robbins (2013). Inhibition and impulsivity: behavioral and neural basis of response control. *Progress in neurobiology 108*, 44–79.

Boureau, Y.-L., P. Sokol-Hessner, and N. D. Daw (2015). Deciding how to decide: Self-control and meta-decision making. *Trends in cognitive sciences 19*(11), 700–710.

Brown, M. R., J. R. Benoit, M. Juhás, R. M. Lebel, M. MacKay, E. Dametto, P. H. Silverstone, F. Dolcos, S. M. Dursun, and A. J. Greenshaw (2015). Neural correlates of high-risk behavior tendencies and impulsivity in an emotional Go/NoGo fMRI task. *Frontiers in systems neuroscience 9*.

Cerigioni, F. (2015). Separating the sheep from the goats: Retrieving preferences when some choices are intuitive. Working paper. Universitat Autonoma de Barcelona.

Cho, I.-K. and A. Matsui (2006). Learning Aspiration in Repeated Games. *Journal of Economic Theory 124*(2), 171–201.

Chung, H.-K., T. Sjöström, H.-J. Lee, Y.-T. Lu, F.-Y. Tsuo, T.-S. Chen, C.-F. Chang, C.-H. Juan, W.-J. Kuo, and C.-Y. Huang (2015). Why do irrelevant alternatives matter? An fMRI-TMS study of context-dependent choice. mimeo.

Corbetta, M., G. Patel, and G. L. Shulman (2008). The reorienting system of the human brain: from environment to theory of mind. *Neuron 58*(3), 306–324.

Corr, P. J. and G. Matthews (2009). *The Cambridge handbook of personality psychology*. Cambridge University Press Cambridge.

Ding, W.-n., J.-h. Sun, Y.-w. Sun, X. Chen, Y. Zhou, Z.-g. Zhuang, L. Li, Y. Zhang, J.-r. Xu, and Y.-s. Du (2014). Trait impulsivity and impaired prefrontal impulse inhibition function in adolescents with internet gaming addiction revealed by a Go/No-Go fMRI study. *Behavioral and Brain Functions 10*(1), 20.

Ferrier, D. (1876). *The functions of the brain*. GP Putnam's Sons.

Fujiwara, E. and H. J. Markowitsch (2006). Brain correlates of binding processes of emotion and memory. In H. Zimmer, A. Mecklinger, and U. Lindenberger (Eds.), *Handbook of Binding and Memory: Perspectives From Cognitive Neuroscience*. Oxford University Press.

Gilboa, I. and D. Schmeidler (2001). *A theory of case-based decisions*. Cambridge University Press.

Glimcher, P. W. and E. Fehr (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan.

Kim, H. F., A. Ghazizadeh, and O. Hikosaka (2015). Dopamine neurons encoding long-term memory of object value for habitual behavior. *Cell 163*(5), 1165–1175.

Kipling, R. (1910). *Rewards and fairies*, Volume 4218. Bernhard Tauchnitz.

Kushner, H. J. and G. G. Yin (1997). *Stochastic Approximation Algorithms and Applications*. Springer-Verlag.

McNay, E. C., T. M. Fries, and P. E. Gold (2000, 03). Decreases in rat extracellular hippocampal glucose concentration associated with cognitive demand during a spatial task. *Proceedings of the National Academy of Sciences of the United States of America 97*(6), 2881–2885.

Schultz, W., P. Dayan, and P. R. Montague (1997). A neural substrate of prediction and reward. *Science 275*(5306), 1593–1599.

Selten, R. (1978). The chain store paradox. *Theory and decision 9*(2), 127–159.

Shomstein, S. (2012). Cognitive functions of the posterior parietal cortex: top-down and bottom-up attentional control. *Frontiers in integrative neuroscience 6*, 38.

Sims, C. A. et al. (2010). Rational inattention and monetary economics. *Handbook of Monetary Economics 3*, 155–181.

Smith, E. E. and J. Jonides (1999). Storage and executive processes in the frontal lobes. *Science 283*(5408), 1657–1661.

Spiegler, R. (2011). *Bounded rationality and industrial organization*. Oxford University Press.

Spieser, L., W. van den Wildenberg, T. Hasbroucq, K. R. Ridderinkhof, and B. Burle (2015). Controlling your impulses: Electrical stimulation of the human supplementary motor complex prevents impulsive errors. *The Journal of Neuroscience 35*(7), 3010–3015.

von Hippel, W. and S. M. Dunlop (2005). Aging, inhibition, and social inappropriateness. *Psychology and aging 20*(3), 519.

Woodford, M. (2012). Prospect theory as efficient perceptual distortion. *The American Economic Review 102*(3), 41–46.