WILLPOWER AND PERSONAL RULES

Roland Bénabou and Jean Tirole

Princeton University IDEI-Toulouse

Journal of Political Economy (August 2004)

I. INTRODUCTION

• Self-control problem: common human tendency to succumb to short-run impulses to seek pleasure or avoid discomfort at the expense of long-run interests (e.g., “I should be saving more”).

• Has recently attracted renewed attention from economists. Often attributed to time-inconsistent preferences (e.g., hyperbolic): the individual's current self “overweighs” the present relative to the future.

• Applied to: addiction, procrastination, self-deception, consumption / savings, portfolio choice, fiscal policy, growth…

• Economics literature either takes it as given that individual is unable to commit to optimal course of action, or else emphasizes the recourse to external commitment devices: avoiding temptation, holding illiquid assets, signing binding contacts, etc.

• This paper: focus on internal commitment mechanisms, or "personal rules,"as emphasized by psychologists: diets, resolutions to jog twice a week, write five page a day, smoke only after meals, always finish what you started, conduct your life with dignity, etc.

• Thus seek to understand genuine self-control or, in the words of Adam Smith, "self-command".

• Question 1: how can such entirely self-imposed rules actually constrain the individual's behavior?

Economically important: distortions emphasized in standard model with no commitment or costly commitment may be overstated.

• Question 2: can personal rules "go too far''? Aim to account for overregulated/compulsive behaviors: workaholism, avariciousness, failure to dissave in old age, anorexia.

Very different set of costs, which have received almost no attention in economics. Likely to also have effects on market and aggregate outcomes.

- Question 3: what are the cognitive underpinnings of self-regulation?

  - Role of recall (or awareness) and attribution (signal-extraction) with respect to one's past feelings, actions, and circumstances (excuses).

  - Role of cognitive rules through which the individual can affect recall and attribution processes: keep a journal, attend therapy / self-help group / confession, rehearse moral or religious principles, etc.

    Key question here also: are these *self-enforcing?*

Key mechanism: self-reputation

• Ainslie (1992) : Lapses treated as precedents, which have adverse impact on future behavior. Raises the stakes on misbehaving today.

• But why? Must be learning something about oneself ⇒

Model of self-reputation about one's own willpower.

• Key role of self-monitoring / self-signaling: inferring one's preferences (and hence likely future decisions) from own past behavior.

• How can this be?

• For actions to have informational value, need some form of imperfect recall / awareness of past feelings and motives.

• A lot of psychological evidence (cited below). Imperfect recall can also be endogenized (BT 2000).

Questions

(1) When can self-control be achieved from concern about self-reputation?

Potential determinants : initial self-confidence, memory, excuses...

(2) Can self-reputation concerns lead to compulsive behavior ("$\beta > 1$"):

workaholism, avariciousness, self-deprivation,..?

(3) One step further: role /sustainability of cognitive rules.

Modeling

Simplest = two periods, $t=1,2$ (discount factor δ).

- Each is divided into two sub-periods each (e.g., morning and afternoon).
- Same self-control problem in each period.

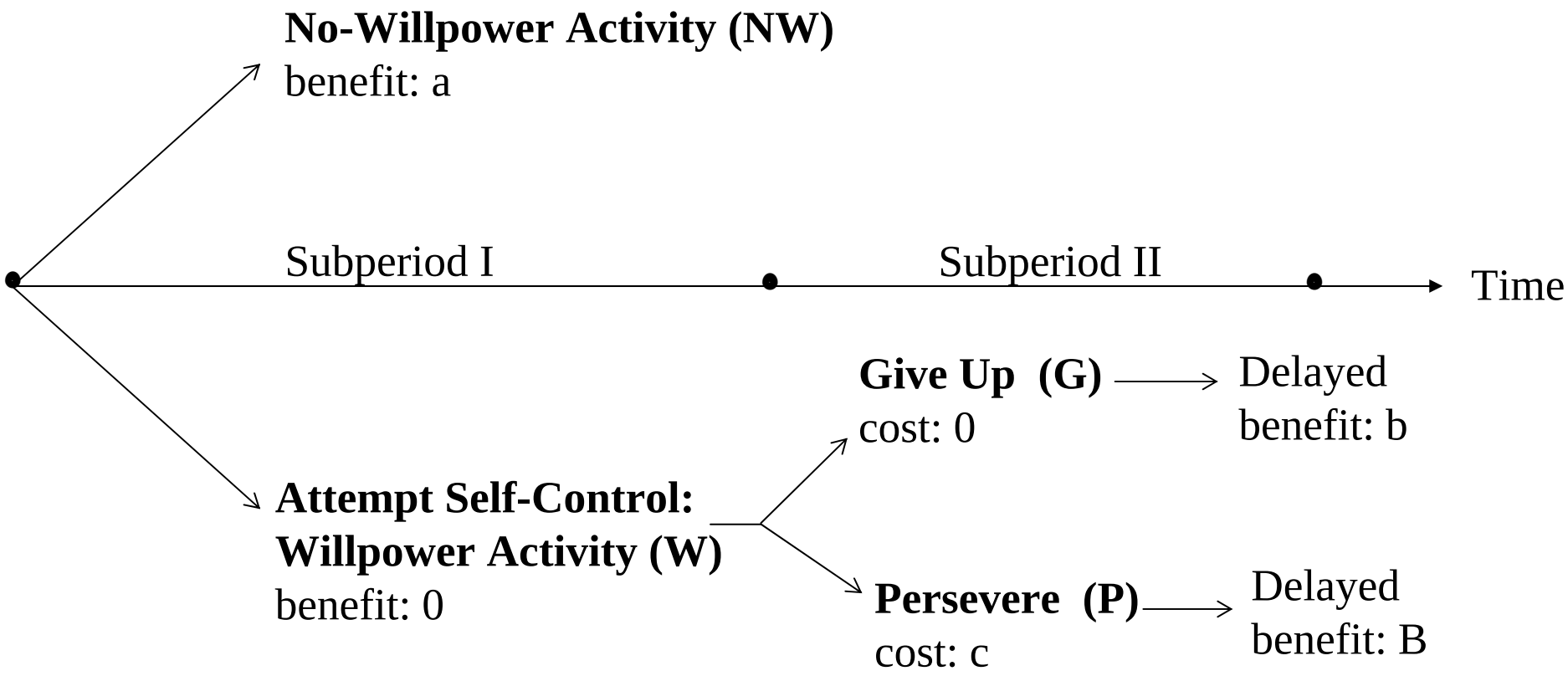


Figure 1: decisions and payoffs in any given period t = 1, 2.

Illustrations

Alcoholic/smoker/overeater

NW:
- binge without restraint
- start drinking in the morning

W :
- attempt self-control: will I refrain or give up anyway (and find excuses)?

Academic

NW:
- take the weekend off
- unambitious research project;

W :
- ambitious project: will I persevere when hardship is encountered?

Social relationship

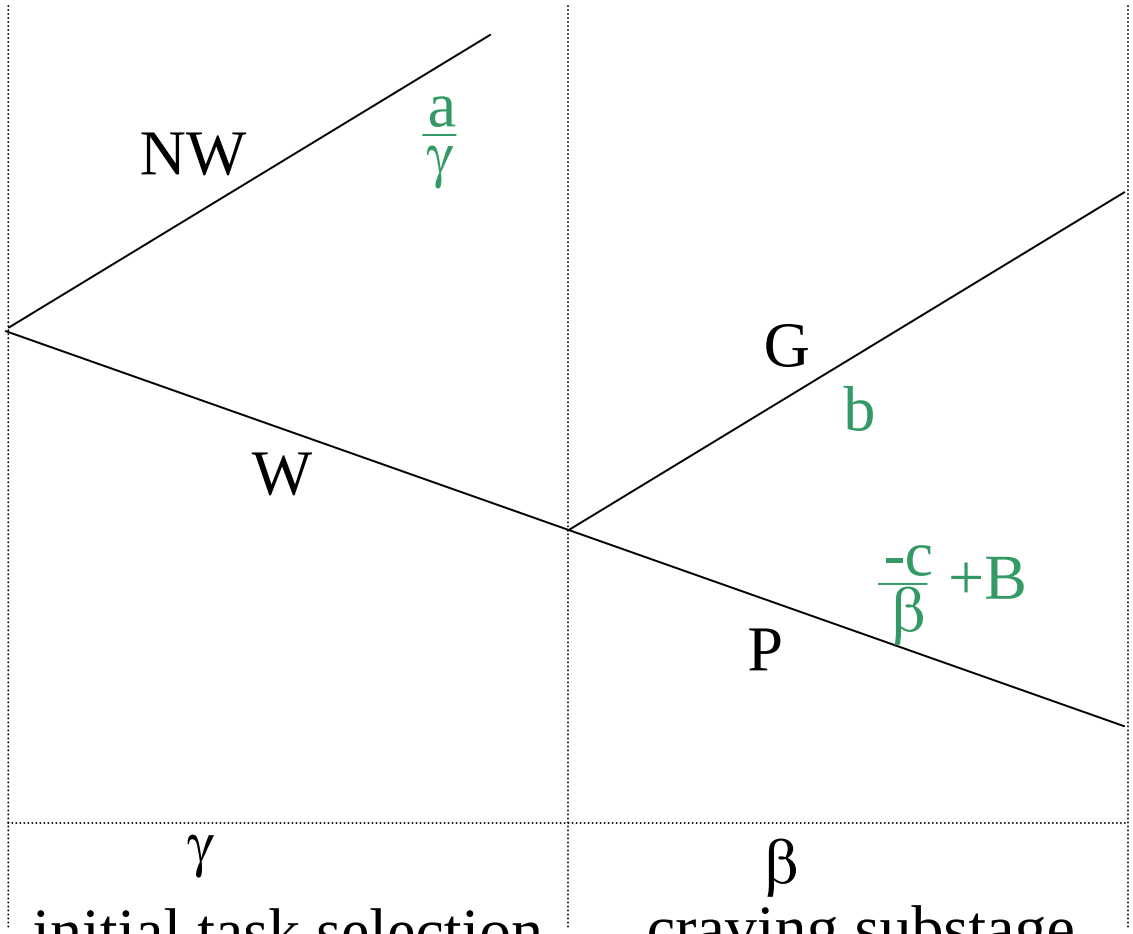
NW:
- isolation (working, surfing web, remain single / short-term interacts.)

W :
- will I do what it takes to build and maintain good interpersonal relationships?

State dependent willpower

Salience of the present: $\gamma \leq 1$, $\beta < 1$ enduring characteristics

(same in both periods).



- Strength of will under stress, β , unknown ex ante
- Rationale: realistic + needed for any self-knowledge / reputation concern
- W will be chosen only if enough confidence in own willpower.

Imperfect knowledge of willpower

Willpower learned only when put to the test:

$$\begin{array}{ccccc} \beta & = & \beta_H & \text{or} & \beta_L \\ \text{prior} & : & \rho_1 & & 1-\rho_1 \end{array}$$

- Since β is a "hot" internal, state $\Rightarrow$ cannot be reliably remembered later on from pure, "cold" introspection. Will have to be inferred from actions, which are closer to "hard" information.

- Consistent with a lot of psychological evidence

- As a result: posterior beliefs at date 2, ρ_2, will be based on (the recall of) past behavior. A "revealed preference'' approach to predicting one's own future choices.

• A lot of evidence that memory about past feelings, motives, "visceral states,'' is:

- *Imperfect:* Kahneman et al. (1997) on experienced/recalled utility, Loewenstein and Schkade (1999) on "hot-cold" empathy gaps.
- *Self-serving:* large literature on motivated recall, self-serving attributions, etc. Modeled in previous paper.

• A lot of evidence also (e.g., Bem (1972), Quattrone-Tversky (1984)) that people:

- make "diagnostics" about their preferences / "the kind of person they are" by observing and interpreting their own behavior.
- conversely, make choices in ways that allow them to maintain certain desirable self-views (e.g., self-respect, identity).

Self-Serving Awareness / Memory

"In science, literature, and folklore, people are famous for … remembering their successes and overlooking their excesses, trumpeting their triumphs and excusing their mistakes, milking their glories and rationalizing their failures, all of which allows them to remain relatively pleased with themselves despite good evidence to the contrary.

Psychologists from Freud to Festinger have described the artful methods by which the human mind ignores, augments, transforms and rearranges information in its unending battle against the affective consequences of negative events."

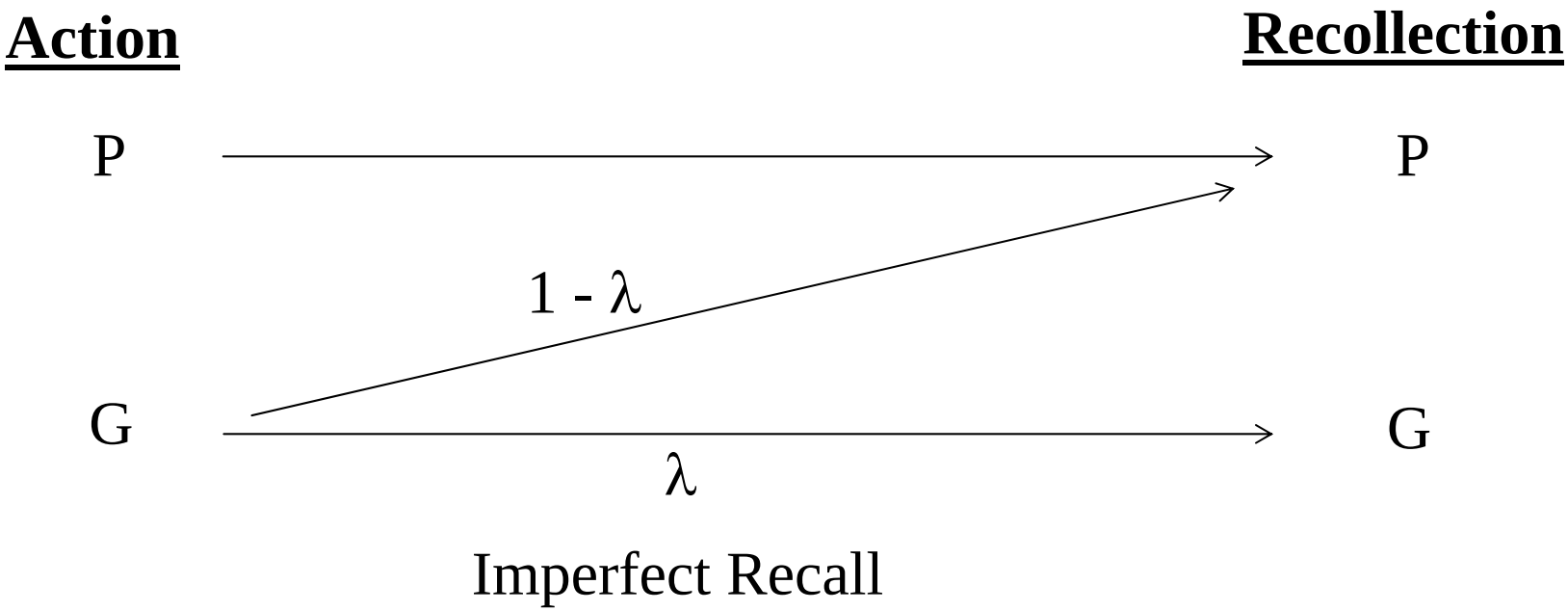
Gilbert et al. (1999)

"We may convince ourselves that we never <u>really</u> loved the ex-spouse who left us for another, but when a friend reminds us of the forty-seven love sonnets that we conveniently failed to remember writing, the jig is up, the fix is spoiled…"

Gilbert et al. (1999)

Imperfect date-2 awareness

Ego-favorable events more likely to be remembered (not crucial)



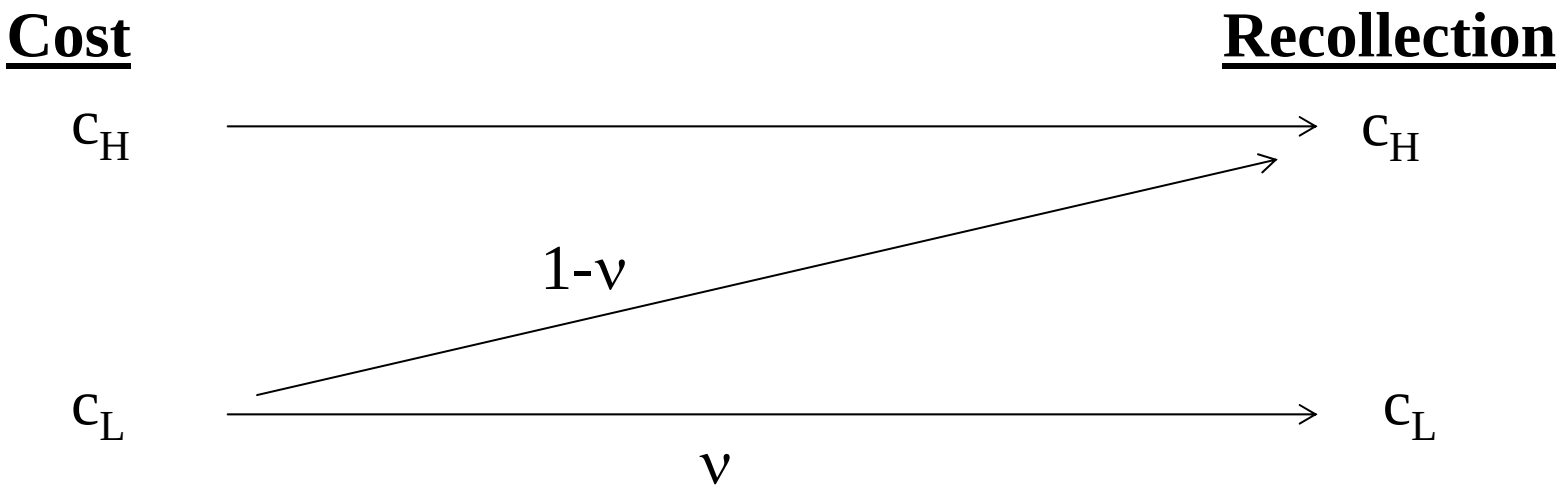
Imperfect Recall

Assumption: Suppose the individual undertakes the *W* activity in period 1. If he perseveres, no lapse will be recalled at date 2. If he gives in to temptation he will remain aware of this lapse only with probability λ. With probability $1 - \lambda$ he will have forgotten" (become unaware of) it, and thus no longer be able to distinguish this state of the world from one where he really held fast.

Cost of resisting impulsion = situational characteristic

$$c = c_L \text{ or } c_H$$

$$\text{probability} : \pi \quad 1-\pi$$



Imperfect Attribution

Assumption: If the cost of effort or craving at time 1 is $c_1 = c_H$, it will never be recalled at date 2 that the task was easy. If $c_1 = c_L$, on the other hand, the individual will recall it only with probability v; with probability $1 - v$ he will no longer not be able to distinguish whether c was equal to c_L or to c_H.

Rules as reputational equilibria

Perfect Bayesian Equilibrium:

Individual
- behaves optimally on the basis of beliefs,
- accounts for impact of choices on future beliefs,
- updates using rational inference (given available information).

Indeed:
- personal rules must be self-enforcing;
- precedents: giving in to temptation today raises the probability that will do the same in the future;
- key role of learning/self-reputation/self-monitoring.

- Rules must be self-enforcing:

"Personal rules are promises to cooperate with the individual's own subsequent motivational states.... They are self-enforcing insofar as the expected value of cooperation exceeds that of defection at the time choices are made. The difference in value can also be regarded as the stake of a private side bet that the person "makes" to precommit his future behavior. It is this stake that gives the will his force." Ainslie (1992)

- How is this achieved? Basic idea is that each decision sets a possible *precedent* for future ones, so that giving in to temptation today raises the probability that one will do the same in the future:

"When particular actions are thus united under a common rule, they are viewed as members of a class of actions subserving one comprehensive end. In this way the will attains a measure of unity."

- Central role of learning / self-reputation:

"But how does a person arrange to choose a series of rewards all at once?...

Assuming that he is familiar with the expectable physical outcomes of his possible choices, the main element of uncertainty will be what he himself will actually choose. In situations where temporary preferences are likely, he is apt to be genuinely ignorant of what his future choices will be. His best information is his knowledge of his past behavior under similar circumstances…

Furthermore, if he has chosen the poorer reward often enough that he knows self-control will be an issue, but not so often as to give up hope that he may choose the richer rewards, his current choice is likely to be what will swing his expectation of future rewards one way or the other."

Ainslie (1992)

III. THE ROLE OF RECALL

- Impact of recall of past lapses (λ)?
- No attribution problem / excuses: one cost, $c_H = c_L = c$
- Assumptions: in full-information context:

(a) Willpower affects self-control: $\frac{c}{\beta_H} < B - b < \frac{c}{\beta_L}.$

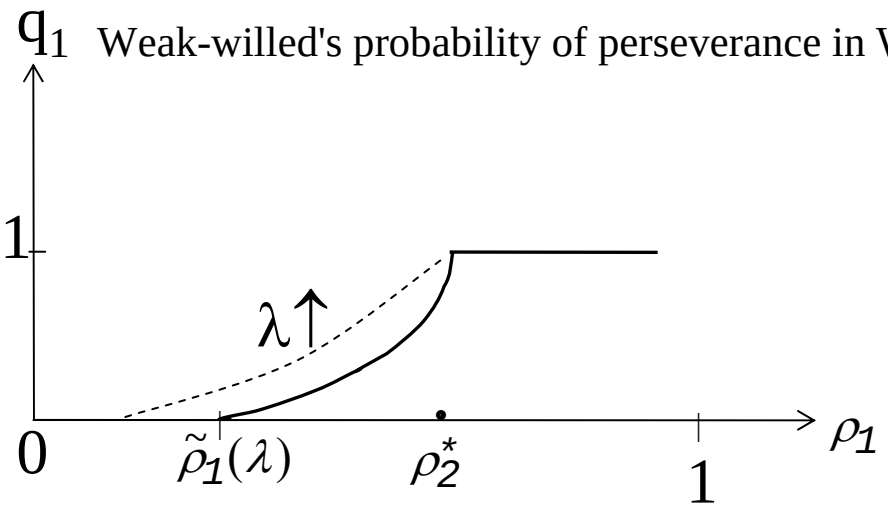
(b) W chosen only if strong-willed: $B - c > \frac{a}{\gamma} > b.$

- Second period: choose W iff $\rho_2 > \rho_2^*$, where

$$\rho_2^*(\mathrm{B} - \mathrm{c}) + (1 - \rho_2^*)b = \frac{a}{\gamma}.$$

Note that when $\gamma < 1$, always *too tempted* to choose *NW* activity.

- First period: unique equilibrium.
  - Strong-willed always perseveres.
  - Weak-willed perseveres if self-reputation is valuable enough (will help with future self-restraint), and lapses are likely to be recalled.
  - Assume $\frac{c}{\beta_L} < B - b + \delta\lambda(b-a)$ (requires $b > a$).



  - W chosen at date 1 iff $\rho_1 \geq \rho_1^*$, where $\rho_1^* < \rho_2^*$.

Determinants of self-control

- Initial self-confidence. As ρ_1 increases:
  - more likely to put willpower to the test,
  - more self-restraint is exercised.
- External control. Suppose: date-1 behavior is externally forced (parents, societal norms, incentives,...), but date-2 behavior is still subject to free will.

Then:
- no impact if initial self-confidence high,
- inferior reputation-building and loss of autonomy if $\rho_1 < \rho_2^*$. Dependence.

- Memory. Lapses less forgettable $\Rightarrow$ more self-restraint. $\lambda = 1$ would be optimal. Role for cognitive rules (self-monitoring, target rehearsal...)

IV. ROLE OF ATTRIBUTION / EXCUSES

- Lapses are remembered: $\lambda = 1$.
- Cost of resisting impulses may not be: $c \in \{ c_L , c_H \}$. Role of ν ?
- Focus on the interesting case:

$$\frac{c_L}{\beta_H} < B - b$$

$$\frac{c}{\beta} > B - b \qquad \text{for all} (c, \beta) \neq (c_L, \beta_H)$$

$\Rightarrow$ date-2 behavior when confronted with W:

	β_H	β_L
c_L	P	G
c_H	G	G

Impulsive rule / laissez-faire.

Potential date-1 behaviors: basic (pure strategy) rules :

	β_H	β_L
c_L	P	G
c_H	G	G

Impulsive (R_0)

	β_H	β_L
c_L	P	P
c_H	P	G

Legalistic (R_1)

	β_H	β_L
c_L	P	P
c_H	G	G

Flexible (R_2)

	β_H	β_L
c_L	P	G
c_H	P	G

Compulsive (R_3)

Mixed rules: R_{ij} = combination of R_i and R_j. For example, R_{02} is:

	β_H	β_L
c_L	P	P/G
c_H	G	G

where P/G denotes a mixed (randomized) strategy. See paper.

Description of equilibrium

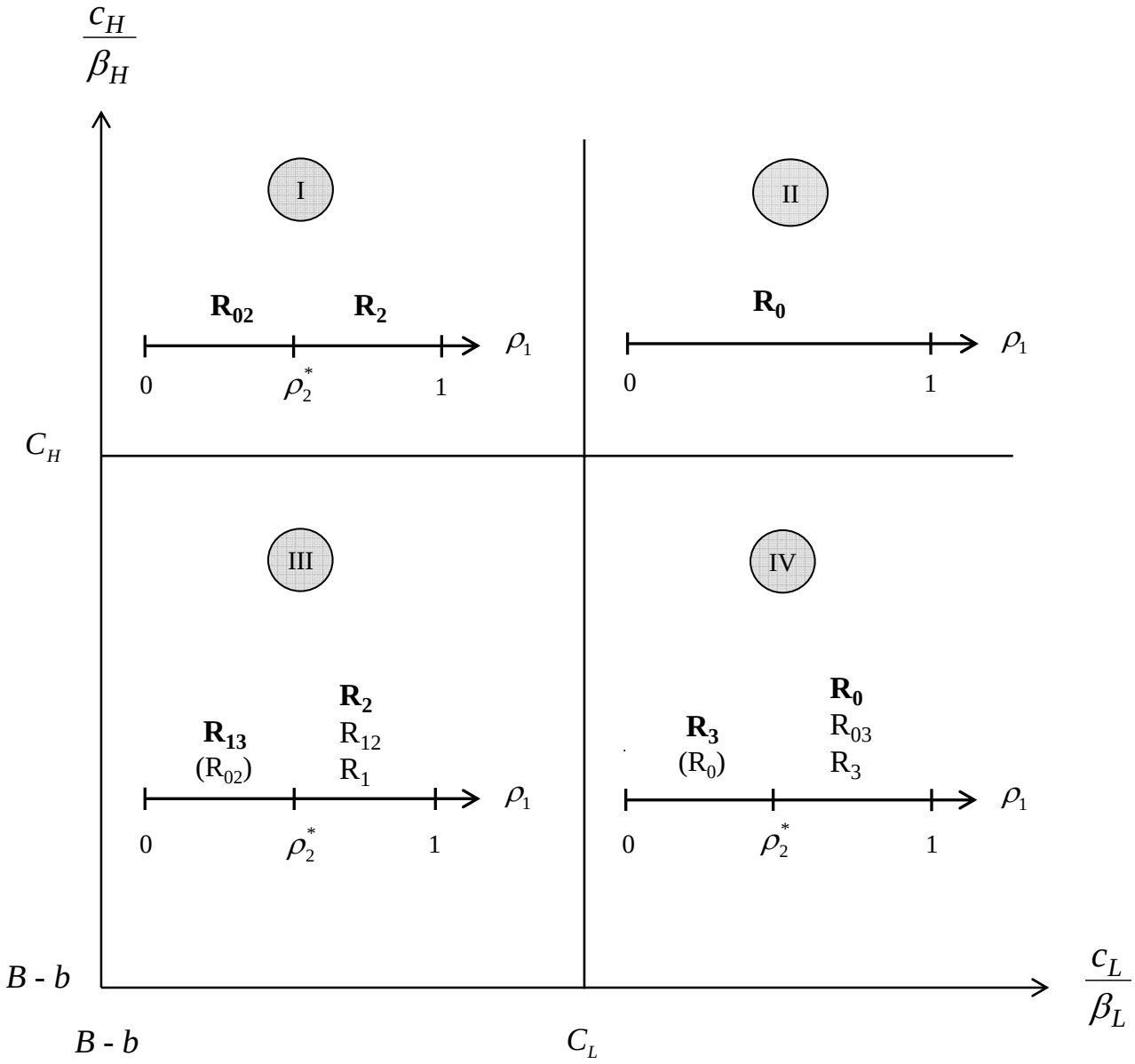
- Maximal gains from self-restraint:

$$\begin{cases} \text{Strong-willed:} & B-b+\delta\left[\pi\left(B-c_L\right)+\left(1-\pi\right)b-a\right]\equiv C_H \\ \text{Weak-willed:} & B-b+\delta\left(b-a\right)\equiv C_L . \end{cases}$$

If $\frac{c}{\beta_i} > C_i$, no self-restraint. In particular, if $\frac{c_H}{\beta_H} > C_H$, $\frac{c_L}{\beta_L} > C_L$ the unique outcome is static, impulsive behavior (R_0).

- Full treatment: see paper. Focus on polar cases $\nu = 1$, $\nu = 0$. Look mostly at two regions exhibiting beneficial self-control and (potentially) harmful compulsiveness. Third one = mixture of these two.

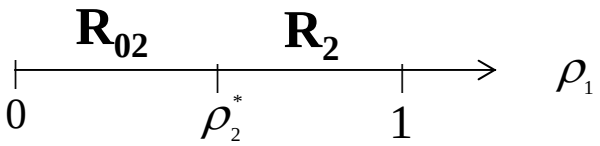
- When multiple equilibria, select Pareto-dominant one (not crucial).


$\frac{c_H}{\beta_H}$
I
II
III
IV
$\mathbf{R_{02}}$
$\mathbf{R_2}$
$\mathbf{R_0}$
$\mathbf{R_{13}}$
(R_{02})
$\mathbf{R_2}$
R_{12}
R_1
$\mathbf{R_3}$
(R_0)
$\mathbf{R_0}$
R_{03}
R_3
ρ_1
0
ρ_2^*
1
C_H
B - b
B - b
C_L
$\frac{c_L}{\beta_L}$

Self-control region

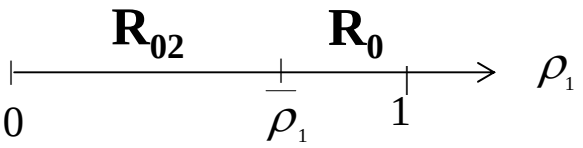
$$\left(\frac{c_L}{\beta_L} < C_L \; ; \quad \frac{c_H}{\beta_H} > C_H \right)$$

- When $c_1 = c_H$, both types yield. When $c_1 = c_L$, strong-willed holds, weak-willed is restrained only by the risk that a lapse for which no plausible excuse is found may lead to complete lack of self-restraint (*NW*) next period.



Partial, flexible S.C Flexible S.C.

$\nu = 1$



Partial, flexible S.C Impulsive

$\nu = 0$

- Perfect inference ($\nu = 1$) ⇒ similar to one-cost case: S.C. increases with ρ_1;
- Less reliable / more malleable inference ⇒ loss of self-restraint; the more self-confident the more so. Thus, S.C. may now decrease with ρ_1;
- Clarity of inference / assessment of excuses (higher ν) benefits individual.

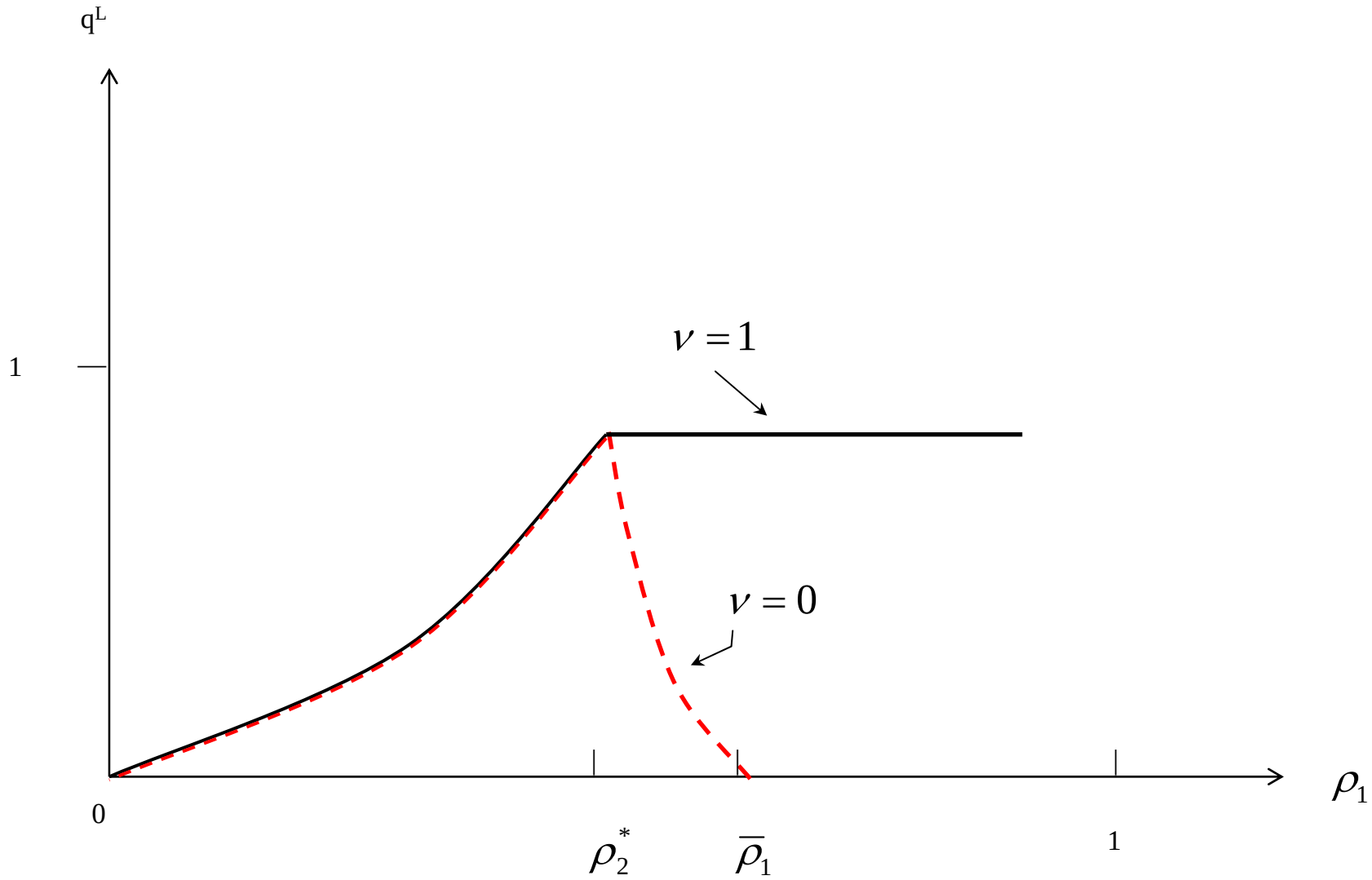
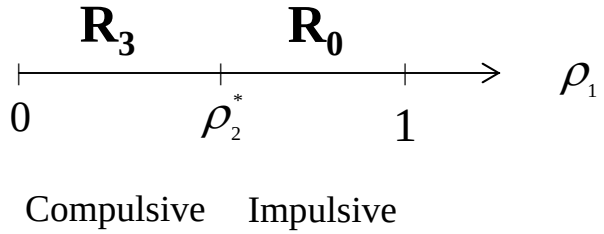


Figure 5a: probability of perseverance by the weak-willed type when $C_1=C_L$: region I, cases ν=1 and ν=0.

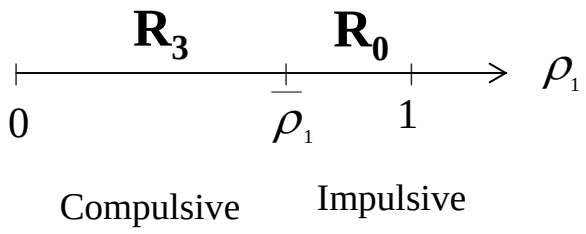
Compulsiveness region

$$\left(\frac{c_L}{\beta_L} > C_L ; \quad \frac{c_H}{\beta_H} < C_H \right)$$

- Weak-willed type always yields. Strong-willed type may hold even when $c_1 = c_H$, due to fear of appearing weak-willed (especially if excuses are not very credible), and causing complete lack of self-restraint (*NW*) next period.



$\nu = 1$



$\nu = 0$

- Compulsiveness more likely if low reputational capital; (also exists for high ρ_1, but dominated by self-forgiving rule R_0);
- Leads to ex-ante welfare loss whenever $B - b < c_H$ (= interesting case);
- More compulsion when excuses are more manipulable / less reliable ($\nu = 0$);
- For low ν, forgetting lapses altogether ($\lambda = 0$) would increase welfare.

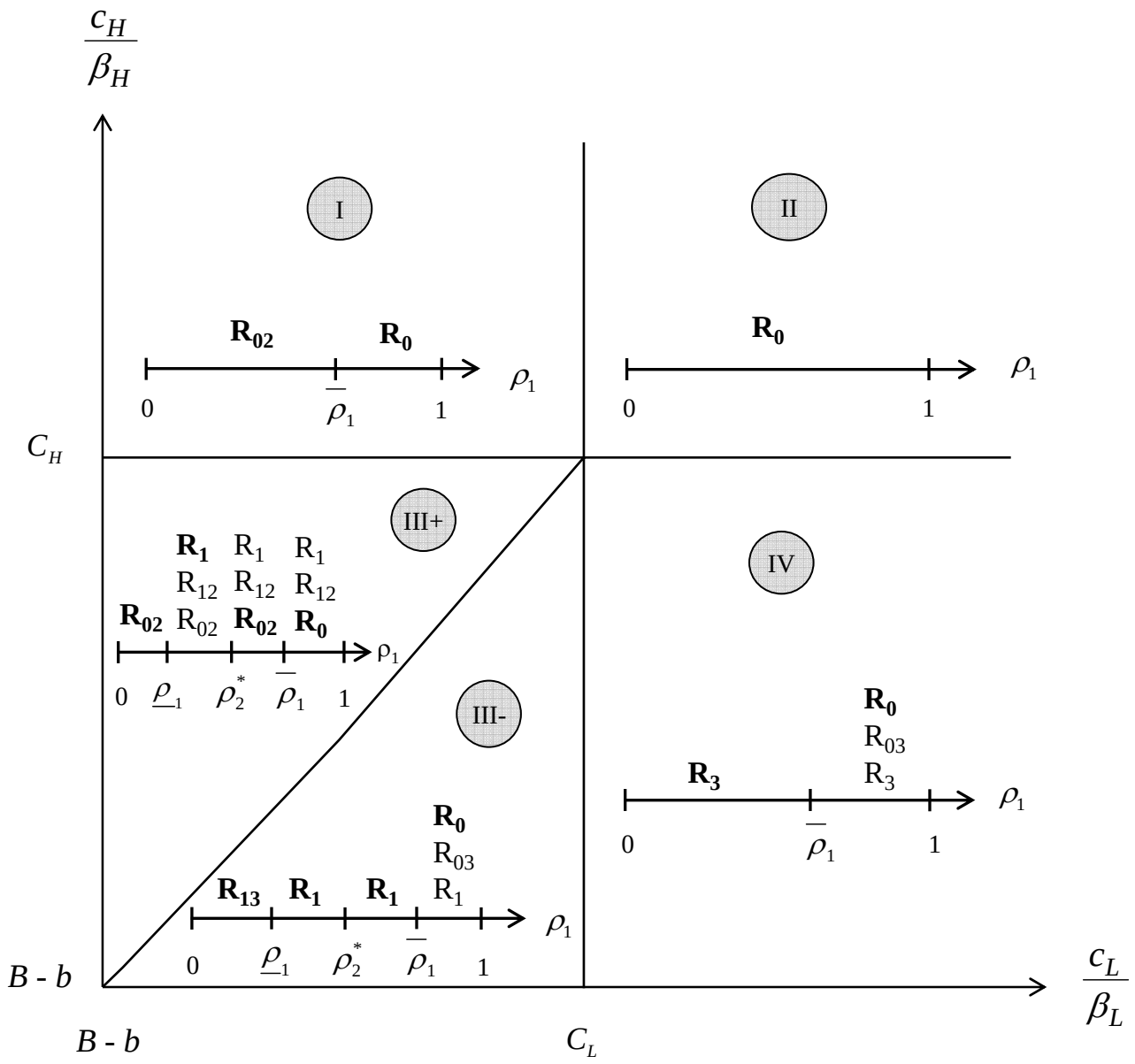
- Standard examples of compulsiveness: miser, workaholic, anorexic, etc. Excessively hard on themselves, “take no excuse”, make no exception.
- Why? Rabid fear of the “slippery slope’’:

“Another version of this theory is that obsessions and compulsions are attempts to compensate for some self-regulatory deficit ...

The obsessional has difficulty in the normal, spontaneous structuring of experience, and therefore tries to compensate by imposing extra structure in the form of boundaries, limits, time markers, and the like...

The quest for such structure, and the excessive adherence to such structure, which have been commonly observed among these individuals, may be a response to the inner sense that they cannot control themselves without those externals aids.’’

(Baumeister et al., 1994).


$\frac{c_H}{\beta_H}$
I
II
$\mathbf{R_{02}}$ $\mathbf{R_0}$
ρ_1
0 $\overline{\rho}_1$ 1
$\mathbf{R_0}$
ρ_1
0 1
C_H
III+
$\mathbf{R_1}$ R_1 R_1
R_{12} R_{12} R_{12}
$\mathbf{R_{02}}$ R_{02} $\mathbf{R_{02}}$ $\mathbf{R_0}$
ρ_1
0 $\underline{\rho}_1$ ρ_2^* $\overline{\rho}_1$ 1
IV
III-
$\mathbf{R_0}$
R_{03}
$\mathbf{R_3}$ R_3
ρ_1
0 $\overline{\rho}_1$ 1
$\mathbf{R_0}$
R_{03}
$\mathbf{R_{13}}$ $\mathbf{R_1}$ $\mathbf{R_1}$ R_1
ρ_1
0 $\underline{\rho}_1$ ρ_2^* $\overline{\rho}_1$ 1
B - b
B - b
C_L
$\frac{c_L}{\beta_L}$

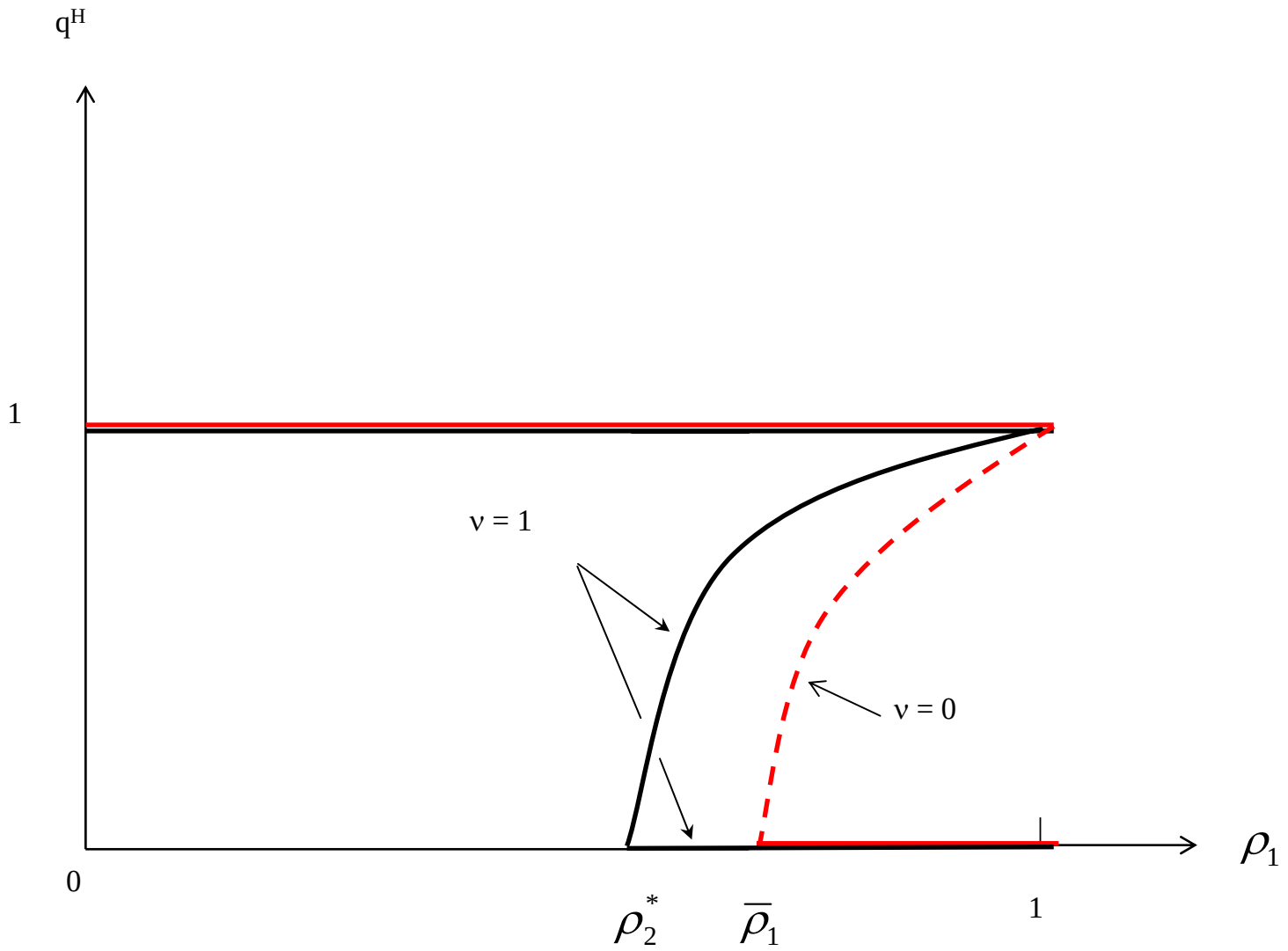


Figure 5b: probability of perseverance by the strong-willed type when $C_1 = C_H$: region IV, cases ν = 1 and ν = 0.

V. COGNITIVE RULES

1. Resolutions

- Specific resolutions vs. broad mandates.
- Moral precepts, religions,... as "solutions" to individual's time inconsistency problem in both personal and social relationships.

2. Why do resolutions matter?

- Affect (λ ,v).
- Redundant channel of activation of lapse-related memory,...

Examples

"Behavior therapists regularly observe that when patients systematically record either impulsive behaviors or avoidances of such behaviors, the occurrence of such behaviors decreases; a practice called self-monitoring."

Ainslie (1992)

"I had during many years followed the Golden Rule, namely, that whenever a published fact, a new observation or thought came across me, which was opposed to my general results, to make a memorandum of without fail and at once; for I had found by experience that such (contrary and thus unwelcome) facts and thoughts were far more apt to escape from memory than favorable ones."

Charles Darwin.

3. Incomplete self-contracts

- Precise rule limits scope for excuses;
- "Bright line" rules as compromise between simplicity and precision.

4. Universality vs. lapse districts

- Willpower is relatively invariant across tasks, activities

  $\Rightarrow$ spillovers in reputation. Similar to "multimarket contact".

- Lapse districts: to limit spillovers from lapses, rehearse information that distribution of cost c is high in some activity (admit helplessness).

V. SUMMARY

(1) Self-restraint more likely if

(a) situation repeated,

(b) lapses more likely to be brought back to awareness,

(c) higher reputational capital? Good for welfare.

(2) Forced choices reduce future self-restraint and may induce dependence.

(3) Situational characteristics (variable costs of resisting impulse) create role for attribution / excuse-making.

(4) Compulsive behavior more likely if

(a) reputational capital is low,

(b) veracity of self-excuses is difficult to ascertain.

(5) Cognitive rules conducive to both self-restraint and compulsiveness.