A COGNITIVE THEORY OF IDENTITY, DIGNITY AND TABOOS

Roland Bénabou - Jean Tirole

Princeton University IDEI-Toulouse

December 2, 2006

INTRODUCTION

- Aim to provide unified treatment of a broad class of phenomena involving *beliefs which people value and "invest in",* with important economic implications: identity, dignity, self-esteem, religion…

  - *Personal:* beliefs about one's deep preferences or "values", abilities, prospects, life after death, ...

  - *Social:* how one fits within / how one values social group (family, firm, peers, culture, nation). How the world works (social mobility, ...)

- Relevant for: cultural integration / immigration, take up of benefits, work/family choices, labor relations, bargaining…

Labor relations / Wage policy

"If you cut the pay of all but the superperformers, you have a big morale problem. Everyone thinks they are a superperformer."

"A pay cut also represents a lack of recognition. This is true of anybody. People never understand and don't want to understand. They don't want to believe that the company is in that much trouble. They live in their own world and make very subjective judgments."

Interviews in Truman Bewley, *Why Wages Don't Fall During a Recession* (1999).

Job search / Take-up of public benefits

But Mr. Rackley refuses to take the [unemployment-insurance] handout. "I was raised to work," he said, "so I swallowed my pride, and now I drive a sod truck." He makes too much money to receive state-financed health care, makes too little to afford his own.

(NYT, October 2006)

Immigration / Integration

"[The] Home Secretary… recommended that minorities speed the process of integration by adopting British "norms of acceptability" and he proposed that newcomers take an oath of allegiance, study British history and culture and embrace "our laws, our values, our institutions."

"Of course it's the wrong thing to be asking of us, said Zahid Hamid, 46, who came here from Pakistan in the early 60's. What a lot of so-called English want us to want is leafy Oxfordshire. But what we want is a job, a decent place to live, safety, a place to educate our children. We want to preserve our separate identities. And remember, we must also maintain the economic link with our original homes. Fourty years later, I am still sending money back."

Britains' Non-Whites feel Un-British, Report says (NYT, 2002).

INTRODUCTION

- Aim to provide unified treatment of a broad class of phenomena involving *beliefs which people value and "invest in",* with important economic implications: identity, dignity, self-esteem, religion.

  - *Personal:* beliefs about one's deep preferences or "values", abilities, prospects.

  - *Social:* how one fits within / how one values social group (family, firm, peers, culture, nation). How society works (e.g., mobility process), life after death…

- Relevant for: cultural integration / immigration, take up of benefits, work/family choices, labor relations, bargaining…

- Many other contexts emphasized by Akerlof and Kranton (2000, 2002).

- Build model from micro level up, making explicit the underlying affective + cognitive mechanisms: a) *self-esteem, anticipatory utility, self-control, b) memory.* Allows unified account for wide range of phenomena / experimental findings.

- Economic applications: hedonic treadmill, taboo tradeoffs, destructive identity, bargaining / scapegoating. Welfare analysis.

- Caveat: less focus so far / less advanced on new experimental / econometric predictions

Outline

Basic framework

- Motivated beliefs: why? Affective and functional motives
- Motivated beliefs: how? Imperfect memory, self-perception / self-signaling

Equilibrium behavior and applications

- Role of uncertainty, salience effect; escalating commitments
- Threats to identity: reaffirmation or compliance

Welfare analysis

- Hedonic treadmill vs. empowerment
- Taboos and "priceless" goods

Multiple dimensions of identity

- Traditional vs. modern identity, dysfunctional behaviors.

Social interactions

- Peer effects, reactions to transgressions
- Bargaining with malleable beliefs: dignity, pride and scapegoating

Conclusion

How to think about self-respect?

- People commonly define or *judge themselves by their own actions*: "I am what I do" (e.g., Adam Smith (1759), Bem (1972), Quattrone and Tversky (1984)).
- Take or avoid actions so as to *maintain or achieve certain views of "who they are":* "keep my self-respect / my dignity", act "as a good Christian," "be true to myself," "maintain my integrity," "stand for my principles," "not betray my values," "be able to live with myself," etc.
- *Actions can only be informative about one's values, character, etc., if those parameters are not directly accessible through introspection / recall.*

How to think about self-respect?

- People commonly define or *judge themselves by their own actions*: "I am what I do" (e.g., Adam Smith (1759); Bem (1972), Quattrone and Tversky (1994)).
- Take or avoid actions so as to *maintain or achieve certain views of "who they are":* "keep my self-respect / my dignity", act "as a good Christian," "be true to myself," "maintain my integrity," "stand for my principles," "not betray my values," "be able to live with myself," etc.
- *Actions can only be informative about one's values, character, etc., if those parameters are not directly accessible through introspection / recall*

Key assumption: individual's true preferences are only episodically accessible to him: limited awareness / retrospective recall of motives and feelings

(e.g., experienced vs. recalled utility, Kahneman et al. 1997; hot/cold gaps in affective forecasting, Loewenstein-Schkade 1999)

⇒ the rest of the time, they will have to be *inferred from past actions*.

⇒ when choosing behavior, will take into account impact on future perception of his own values / type = identity ("what kind of a person would that make me?")

Related literature

Psychology

- *Cognitive dissonnance, self-perception, self-verification.*
  Festinger (1973), Bem (1972), Quattrone and Tversky (1984).

- *Social identity, stereotype threat, in-group / outgroup dynamics:* large literatures.

Economics

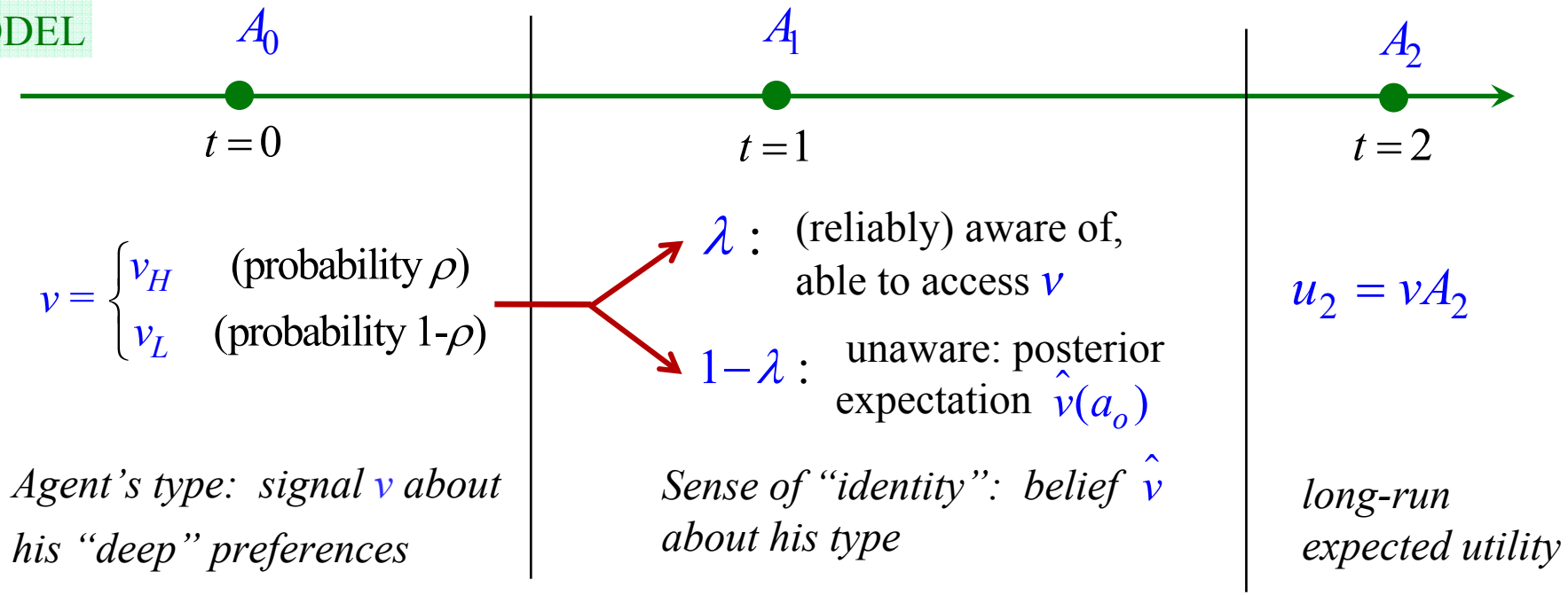
- *Belief distortion / self deception / self image / self signaling*
  Akerlof-Dickens (1982), Carrillo-Mariotti (2000), Bodner-Prelec (2004), Benabou-Tirole (2002), (2004), (2006), Köszegi (2005, Battaglini et al. (2005)…

- *Anticipatory utility*
  Loewenstein (1987), Caplin-Leahy (2001), Landier (2000), Brunnermeier-Parker (2005).

- *Social signaling*
  Bernheim (1996), Austin-Smith and Fryer (2004), …

- *Identity*
  Sen (1985), Akerlof-Kranton (2000, 2004, 2006), Oxoby (2003), Fryer-Jackson (2003), Loury-Fang (2004), Shayo (2004), Horst et al. (2005)…

MODEL



✓ Identity-specific capital: A_t (wealth, human capital, cv, social status, good/bad deeds, family or friends, culture, religion, health; or fixed: gender, race).

✓ Identity-specific activity or investment: $a_t \in \{0,1\} \Rightarrow A_{t+1} = A_t + a_t r_t$

- *How important is A to me in the long run? What are my true values? What kind of a person would investing / not investing in A "make me"?*

✓ Individual's true preference / type v is only episodically accessible to him

The rest of the time, it has to be *inferred from past actions:* $\hat{v}(a_0) = E[v|a_0]$.

1-λ = malleability of beliefs ⇒ allows self-signaling.

What about…

1. Isn't "identity" inherently multidimensional (work/ family, majority/minority culture…)?

a) independent activities and valuations: $(A, v_A; B, v_B ; C, v_C....) \Rightarrow$ same;

b) tradeoff between two dimensions, uncertainty over how much cares about one *relative to the other:*

[- Can invest in either A = work ($a = 1$) or B = family ($a = 0 = 1 - b$). Returns r_A, r_B, salience s_A, s_B, same for other parameters.

- Relative preference shock: $v_A = v_A + v/2$, $v_B = v_B - v/2$, where $v = \varepsilon > 0$ (prob: ρ) or $v = -\varepsilon$ (prob: $1-\rho$)

$\Rightarrow$ same, with $A' = (A - B)$, $r' = (r_A - r_B)$, $s' = (s_A - s_B)$, etc.]

2. Isn't "identity" always socially determined?

- social environment (starting with family) may be key determinant of endowments A, B, … (wealth, education, race, culture) as well prior beliefs ρ (religion, politics);

- may also affect information flows (λ), updating of ρ, salience (s), etc.

- could also affect payoffs (r) and costs of investment: standard externalities.

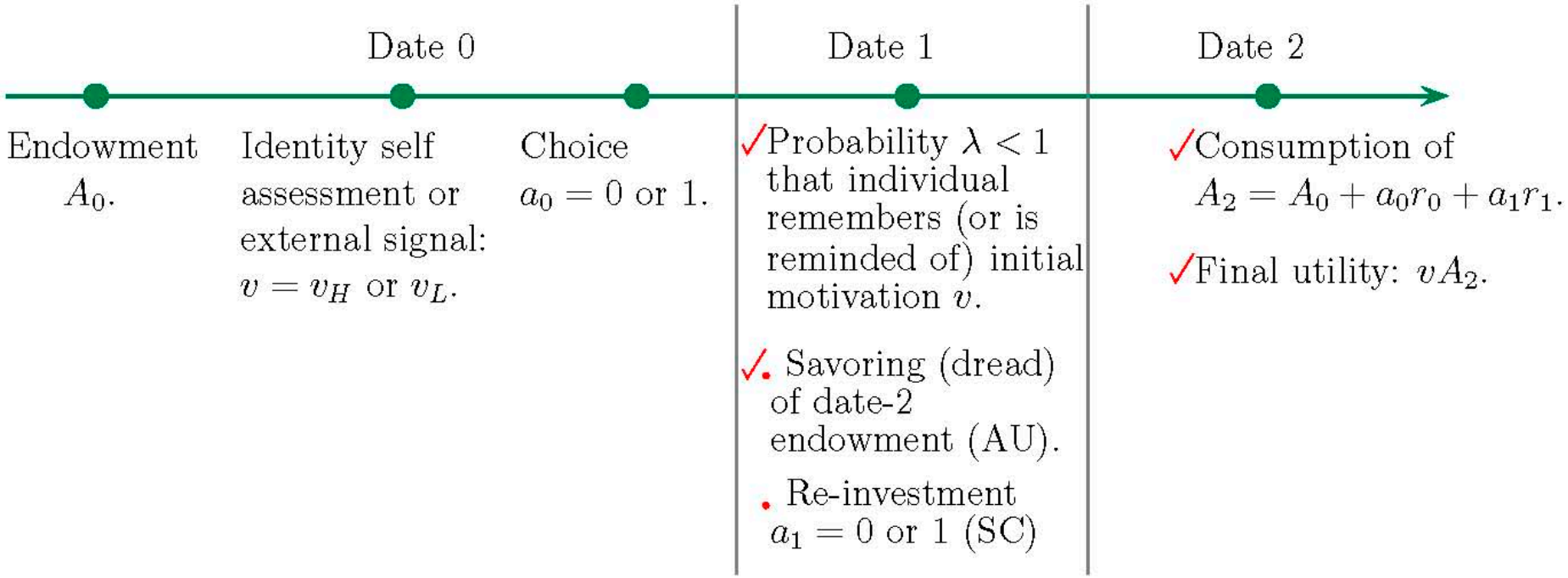
3. Are there alternatives to people having imperfect recall of their motives and feelings?

a) Conscious vs. subconscious knowledge (Bodner-Prelec 2004): the agent and the inner judge, ego and superego, etc., are contemporaneous. Can think of it as case of "instantaneous" forgetting and signaling.

b) Intergenerational transmission of beliefs: children form their "values" ($\hat{v}$) in part from what they see their parents do, or from what the parents force them to do. Can think of it as generation-interval forgetting and signaling.

c) Also care about perceptions of / signaling to others –real or imagined (Adam Smith).

⇒ *Different interpretations or even different phenomena, but all formally the same.*



Date zero payoff + *Continuation value*

$U(v, A_0, a_0)$ $\quad$ $V(v, \hat{v}, A_1)$

Self-view / beliefs at $t = 1$ affect *expected welfare* and *behavior* ⇒

At $t = 0$, incentive to distort action so as to *manipulate later beliefs*

} key tradeoff

($U_{13} \geq 0$, $U_{23} \geq 0$ ← *sorting conditions* → $V_2 > 0$, $V_{12} \geq 0$, $V_{13} > 0$)

Demand for Beliefs 1: Anticipatory Utility or Self-Esteem

- Single investment decision, at $t = 0$ only $\Rightarrow A_2 = A_1$
- Signal / type $v = v_L, v_H$ at date 0 → value of holding belief $\hat{v}$ and stock A_1 in period 1 is

$$V(v, \hat{v}, A_1) \equiv \left(\delta_1 s_1 \hat{v} + \delta_2 v\right) A_1$$

Demand for Beliefs 2: Willpower / Self Control

- Decisions now at $t = 0$ and at $t = 1$ (reinvestment, persistence, etc.). Investing / acting at date 1 is ex-ante efficient for both types, but subject to willpower shock β_1

$$V(v, \hat{v}, A_1) = \delta_2 v A_1 + \left(\delta_2 v r_1 - \delta_1 c_1\right) \times \Pr\left[\beta_1 \delta_2 \hat{v} r_1 \geq \delta_1 c_1\right]$$

- Having a stronger identity helps in making choices, persevering, resisting temptations.

Demand for Beliefs 1: Anticipatory Utility or Self-Esteem

- Single investment decision, at $t = 0$ only $\Rightarrow A_2 = A_1$
- Signal / type $v = v_L, v_H$ at date 0 → value of holding belief $\hat{v}$ and stock A_1 in period 1 is

$$V(v, \hat{v}, A_1) \equiv \left(\delta_1 s_1 \hat{v} + \delta_2 v\right) A_1$$

Demand for Beliefs 2: Willpower / Self Control

- Decisions now at $t = 0$ and at $t = 1$ (reinvestment, persistence). Investing / acting at date 1 is ex-ante efficient for both types, but subject to willpower shock β_1

$$V(v, \hat{v}, A_1) = \delta_2 v A_1 + \left(\delta_2 v r_1 - \delta_1 c_1\right) \times \Pr\left[\beta_1 \delta_2 \hat{v} r_1 \geq \delta_1 c_1\right]$$

- Stronger identity helps in making choices, persevering, resisting temptations
- Both cases: date-0 instantaneous payoffs $U \sim$ effective *cost of investment* $c_0^H \leq c_0^L$.
- Overall welfare (basic case): $W \equiv E[U + V]$

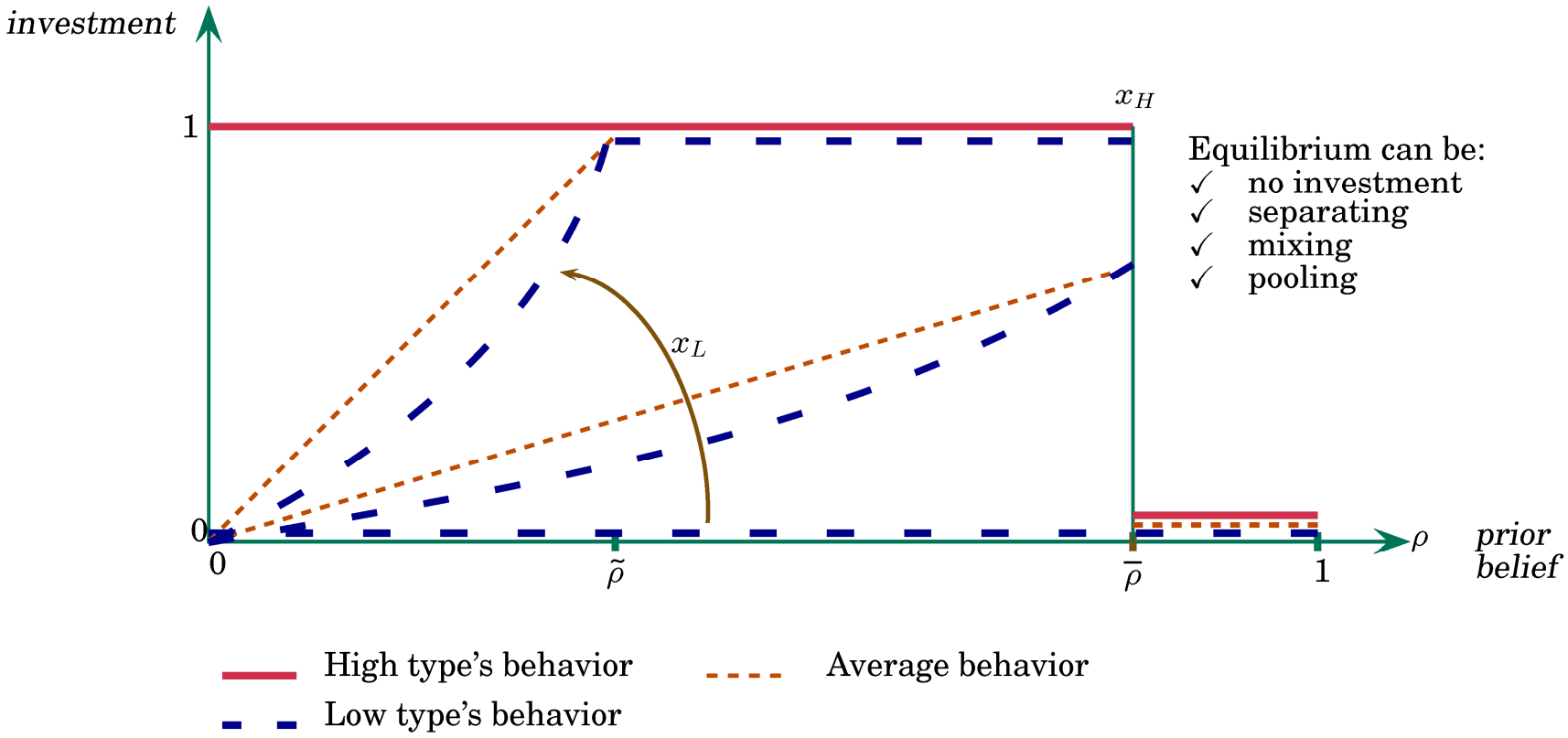
Can mix $AU + SC$

EQUILIBRIUM

• Behavior: probabilities x_H, x_L that someone with signal of being high / low valuation type invests at $t = 0$. Optimally chosen, given anticipated costs and benefits, including hedonic / and or instrumental value of the self-image / identity that will result at $t = 1$.

• *Beliefs:* in drawing (self-) inferences, individuals are sophisticated / Bayesian.

Proposition 1. *There exists a unique (monotonic, undominated) equilibrium, such that:*



Proposition 2. *(1) An individual invests more in identity (x_L and/or x_H rise),*

(i) the more malleable his beliefs (lower λ),

(ii) the more salient the identity (higher s_1) under anticipatory utility

(iii) the higher his identity-specific capital (A_0) under anticipatory utility

(2) The strength of the initial belief or identity ρ has a non-monotonic (hump-shaped) effect on average investment.

⇒ Applications

✓ (Objective) information-poor environments and imperfectly known preferences increase identity investments (new immigrants, converts, born-again, adolescents, etc.)

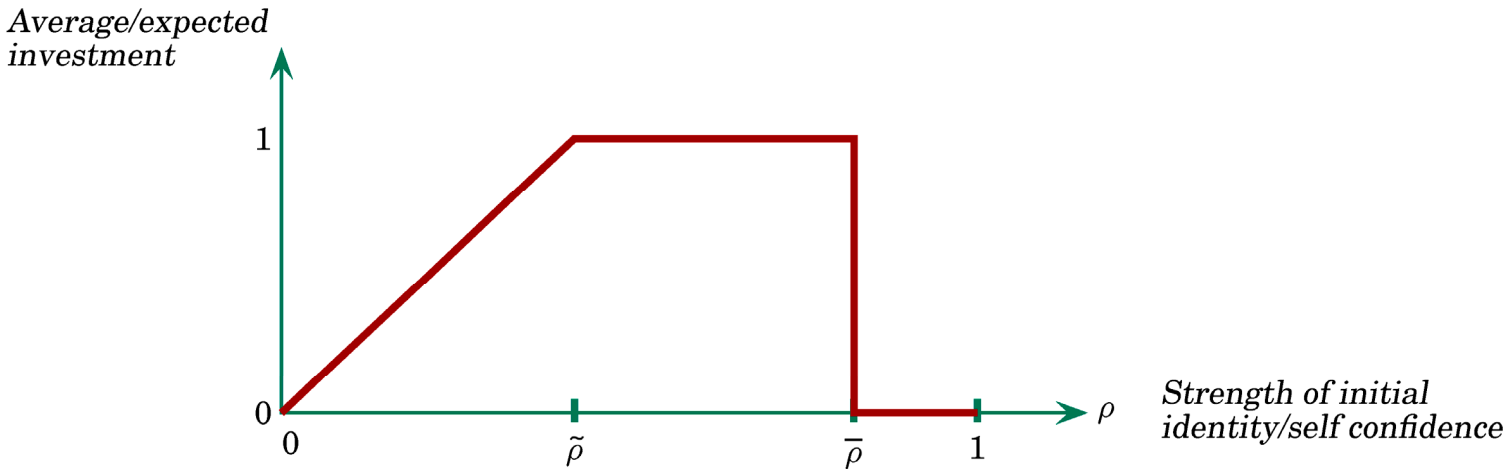
✓ Manipulating salience of a valued identity leads to identity- affirming choices for consumption, investment, etc. (e.g., LeBoeuf and Shafir 2004, Benjamin et al. 2006).

✓ *Escalating commitments:* the more A_0 you have, the more important it becomes to think that it is (ultimately) valuable. The way to "demonstrate" such beliefs is to *invest further:* "stay the course". Raises A_1 even more, etc.

⇒ People who "define themselves" by their work, culture, religion, etc. Managers, farmers who keep "throwing good money after bad". Psy literature on self-justification (Staw 1976).

Applications (continued)

✓ *Identity threats (lowered ρ):* hill-shaped behavior ⇒ whether "fight" or "concede" depends on prior level and uncertainty over identity: where was ρ relative to $\tilde{\rho}$ *and* $\bar{\rho}$.



✓ Threats to strongly held identity ⇒ strong opposing responses, meant to "repair" the damaged beliefs: religious identity (e.g. Danish cartoons), sexual identity (Maas et al. 2003), good-person identity ("transgression-compliance" effect, Carlsmith-Gross 1969).

✓ More subtle challenges to / affirmations of relatively fragile or unfamiliar identity lead to confirmatory responses: "foot in the door" effect (e.g., DeJong 1979), academic "stereotype threat" (e.g.. Steele and Aaronson 1995).

Welfare analysis: Is Identity Good for You?

1. Anticipatory utility / self-esteem : treadmill effect!

Proposition 3. *In the AU version,*

(1) A greater malleability of beliefs (1-λ) always reduces ex ante welfare.

(2) An increase in (per se valuable) identity-specific capital A_0 can also reduce welfare.

• Intuition: average reputation is fixed => signaling-motivated investments just lead to deadweight loss (obvious when stock is immutable: $r_0 = 0$):

$$W = \rho x_H \left[(\delta_1 s_1 + \delta_2) \; v_H r_0 - c_0^H \right] + (1-\rho) \; x_L \left[(\delta_1 s_1 + \delta_2) \; v_L r_0 - c_0^L \right] + \left[s_0 + \delta_1 s_1 + \delta_2 \right] \; \bar{v} A_0 .$$

➢ Hedonic treadmill: higher wealth, social / professional status, etc., need not increase life satisfaction that much, may even reduce it, precisely because *trigger self-defeating pursuit* of the belief that these assets will bring long-run happiness!

2. Self control / time inconsistency

Proposition 4. *In the SC version, a greater malleability of beliefs (1-λ) can raise welfare, by enhancing motivation and improving choices at t = 0 and / or t = 1.*

⇒ Similar positive implications, very different normative ones

Taboos and Sacred Values

- Economics: all goods are *fungible* or "secular", i.e subject to trade-offs at some price (market or shadow).

- All societies, religions, cultures hold, or at least declare, certain things to be *"priceless" or "sacred":* life, liberty, justice, honor, love, friendship, one's children, faith, etc.

- Many markets banned because viewed as *"contrary to human dignity",* harmful by their mere existence. "Commodification" of life, death, sexuality, human organs, genes, environment, morality, etc., as this would "debase" higher ideals. *"To compare is to destroy"* (Fiske and Tetlock 1997). But *destroy what, how?*

- Taboos and sacred values = *upholding certain beliefs (true or illusory),* deemed vital for the individual or for society, concerning things one "would never do" and the "incommensurable" value of certain goods.

- For either *anticipatory-utility* (including prospects of *afterlife*) or *self-discipline* motives, may want to be optimistic about value v of freedom, bodily integrity, non-addiction, relationship to a person (child, spouse, friend) or to a more abstract entity (country, religion) → continuation value function $V(v, \hat{v}, A_1)$.

- At $t = 0$, agent can find out the "sellout" price p at which he could exchange one unit of A_0 against money or other material goods of known consumption value. Ex ante,

$$p = p_H \ (probability\ z) \quad or \quad p = p_L \ (probability\ 1 - z)$$

- Tradeoff p may be learned by checking price offered on a formal or informal market (for switching political loyalties, selling one's vote, organ or children; for prostitution, fraud, crime, etc.) or by simply engaging in deliberate, "coldhearted" calculations about the costs and benefits of different courses of action.

- Will later recall whether or not *entertained* the possibility of a transaction, *evaluated* whether maintaining his identity, dignity, etc., was "worth it" or not ⇒ draw from this the appropriate *inferences about where his "true values" lie.*

- Will *uphold the taboo* against finding out p if foregone option value is not too high:

$$\mathbf{V}(v, \hat{v}(1), A_0) - \mathbf{V}(v, \hat{v}(0), A_0 - z) \geq z p_H \ .$$

- Positive results: how sacred values arise and are sustained, by all or by some; how *taboo-breaking* by others can lead to reaffirmation or collapse.

- Normative results: welfare effect (at individual level) of taboos depends critically on whether they reflect *"mental consumption"* or *self-discipline* motives.

Multiple Identities

✓ *Conflicts:* work / family, own / dominant culture (immigrants, minorities), wealthy / liberal, college-bound / neighborhood (~ Austen-Smith and Fryer 2005)

✓ *Complementarity / clusters:*

- Lamont (2002): "caring self" (generosity, solidarity, family, friends, …) vs. "disciplined self" (work ethic, willpower, responsibilities,…).
- Kunda (2002), Nisbett (2003): independent self (West) vs. interdependent self (East).

➢ *Sources of interaction among identities*

- Resource rivalry: e.g., can only invest in A or B (e.g., time constraint)
- Consumption rivalry: will eventually consume only A or B good (specialization)
- Affiliation: v_A and v_B positively or negatively correlated

Multiple identities and dysfunctional behaviors

- *Modern identity B:*

- Known value v_B: easy to quantify, "secure". Monetary benefits of new job, assimilating into dominant culture, etc.

- Investment is risky: $b_0 \equiv 1$, cost $c_B \rightarrow$ return r_B with probability z (success) or 0 with probability $1-z$ (failure). Education, new skills, new location or social networks.

- *Traditional identity A:*

  - No further investment (for simplicity): $a_0 \equiv 0$. Thus A represents fixed trait like ethnicity, long-held skills, connections to "the old country", etc.

  - "Insecure": hedonic value more subjective, less quantifiable: durability and importance of personal commitments, long-run utility from family, culture, religion, morals: $v_A = v_H$ *or* v_L, with probabilities ρ and $1 - \rho$.

- Date 0: signal $v_A \Rightarrow$ date 1: awareness probability: $\lambda < 1$.

- Date 2:
  - aware of v_A (for simplicity)
  - must choose between *consuming either A or B* (consumption rivalry): in what sector will work? In which country / culture will retire / raise children?

Simple example with Anticipatory Utility (point is more general)

- At t =1, will choose to consume B only if successfully invested in it; A = fallback

$$v_B B_0 < v_L A_0 < v_H A_0 < v_B \left(B_0 + r_B \right)$$

- Investing in *B* is sufficiently productive that, *absent identity concerns* (or, with non-malleable beliefs), *everyone would do it*, even those with high value for A

$$z\left(\delta_1 s_1 + \delta_2\right)\left[v_B\left(B_0 + r_B\right) - v_H A_0\right] > c_B$$

Also implies that investing is dominant strategy if no consumption rivalry at t = 1.

➢ Nonetheless, *neither type invests* (unique equilibrium), if

$$\underbrace{z\left(\delta_1 s_1 + \delta_2\right)\left[v_B\left(B_0 + r_B\right) - v_L A_0\right]}_{\textit{economic return}} - \underbrace{(1-z)\delta_1 s_1 (1-\lambda)\left[\bar{v} - v_L\right]}_{\textit{identity loss}} A_0 < c_B$$

Intuition: non-investment in B is similar to investment in A in the basic model.

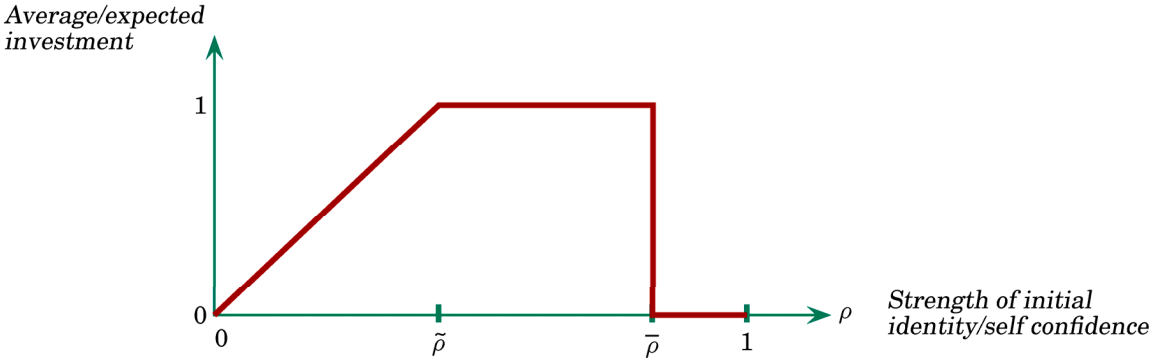
⇒ Individual may not invest in B even when it is efficient to do so (for both types), for fear that it *would convey bad news about* v_A . (~ Austen-Smith and Fryer 2005).

⇒ Applications

- Resistance / hostility to technical change / globalization: alter relative payoffs of traditional / modern sector, but transition requires risky investments .
- Immigrant / expatriate concerned about losing (or not passing on) his culture / religion: may resist assimilation / forego valuable investments in local capital: human, social, housing, retirement assets. Will resent having to take oath of allegiance, dress codes,….
- *Destructive identity:* "not investing in B" can also mean actually disinvesting, by destroying some B capital: same model with $c_B < 0$. French riots: youths destroying schools, day care centers, pharmacies, cars, in own community.
- People can tip from (optimally) investing in B to (self-defeatingly) destroying B if perceive reduced chance of success z (e.g., via education) or lower payoffs r_B (e.g., labor market discrimination), or if salience s_1 of alternative identity in which they "seek refuge" is raised by ideological manipulation or media attention.

Peer effects and responses to transgressions

- ✓ Direct preference spillover: on v , r or c. Familiar, will abstract from it.
- ✓ Cognitive channel: v_1 and v_2 correlated ⇒ individuals' self-view / world view is affected by observing others' behavior. (~ Battaglini et al. 2005).



➢ *Response to in-group member* j 's identity-consistent ($a_j = 1$) or identity-inconsistent ($a_j = 0$) behavior: can directly apply previous results on changes in ρ.

➢ *More similar reference group:* same as mean-preserving spread in ρ (intuition: j's behavior is more informative). At initially high levels, tends to raise identity investment ($a_i \nearrow$) to respond to challenge; at initially low levels, tends to further "sap morale" ($a_i \searrow$) .

➢ Response to transgressions can take different forms: re-investment, exclusion of deviators (lowers λ: out of sight, out of mind), harassment / punishments.

Dignity and scapegoating in bargaining or group conflict

- Pride, dignity, wishful thinking lead people or groups to *walk away* from "reasonable" offers, try to *shift blame* for failure onto others, *destroy surplus,* take refuge in political utopias ⇒ costly delays, impasses and conflicts. Trials, divorces, strikes, scapegoating of minorities in hard times, wars.

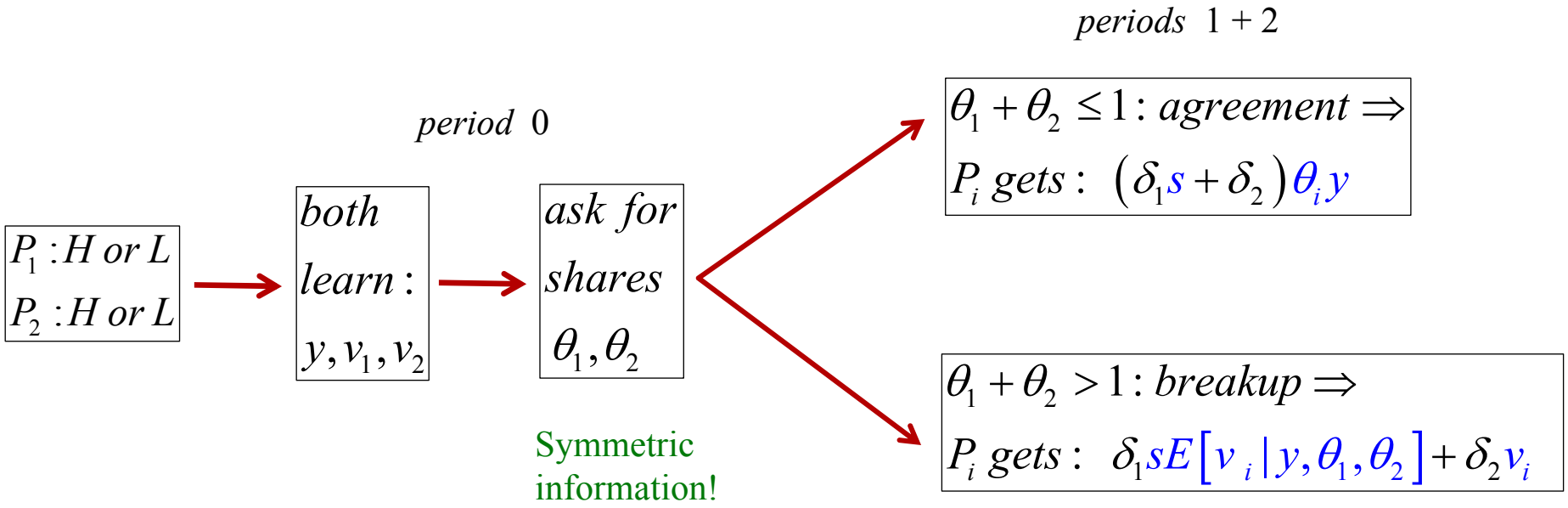
- Importance of *belief distortion* in those phenomena: field observers (Bewley 1999) + experiments, e.g. Babcock, Loewenstein et al. (1995): subjects in bargaining situations with *common knowledge* spontaneously generate, through *self-serving processing and recall of the evidence,* divergent fairness judgments and deluded predictions of outcomes; those, in turn, lead to costly delays and failures to agree.

⇒A simple model of self-serving biases and Coasian failures

- Two-member "partnership" (spouses, capital-labor, ethnic majority/minority) produces joint output y, level of which can be good or bad*:* $y = y_G$ *or* $y = y_B$.

- Each side has type *H* or *L*: ability, motivation, etc. Technology such that low output means *at least one member* is low type (perhaps both): $HH \rightarrow y_G$*;* HL *or* $LL \rightarrow y_B$.

Bargaining with malleable beliefs

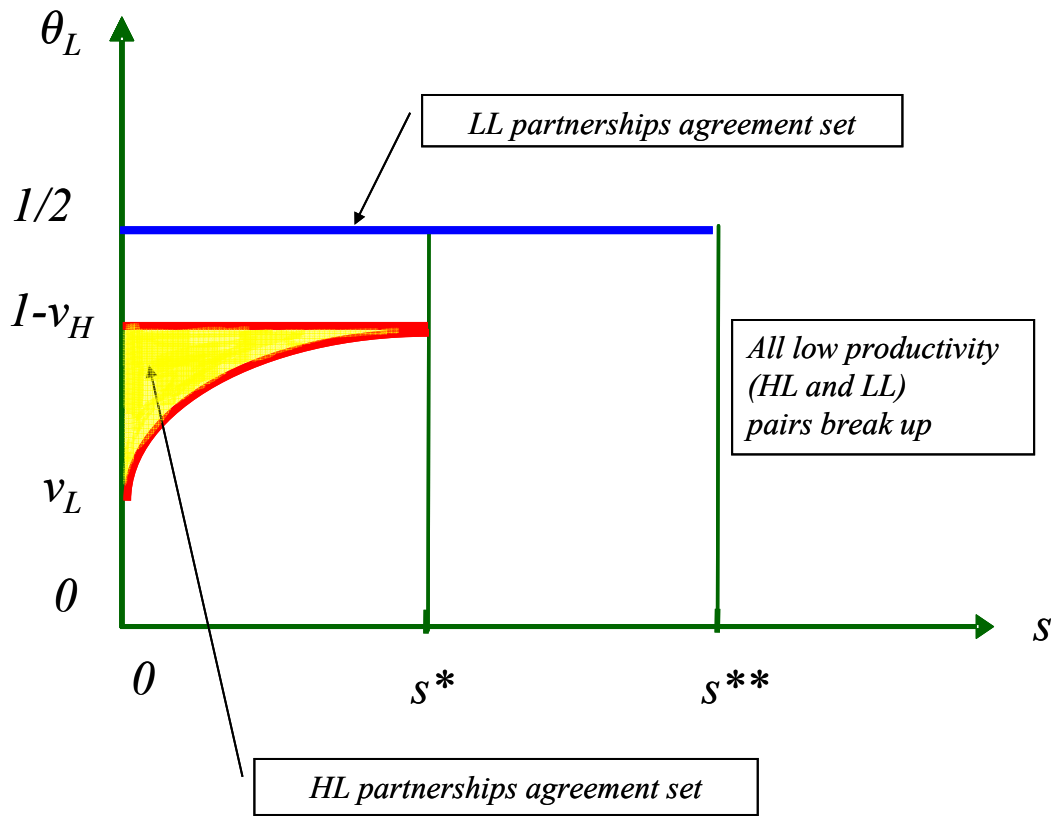


- Joint output y , offers θ_i = hard data, easy to recall / verify. *Individual contributions* or values v_i = *soft information, malleable beliefs* ($\lambda = 0$, for simplicity).
- *Anticipatory utility* / *self-esteem, pride*: same s for both players

⇒ Incentive to refuse low offers / demand high share / walk away, to preserve or achieve the view that one is an H type / the other side is to blame for the low output.

- Look for pure strategy, symmetric equilibrium: shares $\theta_L^* < 1/2 < \theta_H^*$ for L and H in an *unbalanced team,* share 1/2 for both in a *balanced* one (HH or LL).
- Belief restrictions off the equilibrium path (more)

Proposition 8. *All dissolutions are inefficient. Yet, there exist s* and s** such that:*



"If you cut the pay of all but the superperformers, you have a big morale problem. Everyone thinks they are a superperformer."

"A pay cut also represents a lack of recognition. This is true of anybody. People never understand and don't want to understand. They don't want to believe that the company is in that much trouble. They live in their own world and make very subjective judgments."

Interviews in Truman Bewley, *Why Wages Don't Fall During a Recession* (1999).

Conclusion

• Simple model for analyzing broad set of beliefs which people value and invest in: identity, dignity, "a better tomorrow", religion, etc.

• *Unified account* for a number of findings from psychology: salience effects, escalating commitments, responses to identity / stereotype threats…

• Economic implications: excessive persistence / specialization, hedonic treadmill, destructive identity, taboos against explicit prices, failures of Coasian agreements.

Avenues for future work:

• Endowment effects.

• More on taboo tradeoffs and "sacred" values.

• Bargaining / distributive conflict with malleable beliefs: applications to contracts, organizations, political economy.

Example 3: Wishful Thinking and Procrastination (AU + SC)

- When does desire to indulge in pleasant beliefs / avoid unpleasant thoughts aggravate self-control problem, and when does it alleviate it? Complementarity vs. substitutability.

- Combine AU + SC and allow investment to have type-dependent returns, $r_t(v)$, $v = v_H, v_L$

$\Rightarrow$ contribution to long-run welfare vA_2 is $z_t(v) = v \cdot r_t(v)$.

➢ Agent at $t = 1$ now invests when $\beta_1 (\delta_1 s_1 + \delta_2 v) z_1(\hat{v}) \geq \delta_1 c_1 \Rightarrow$

$$V(v, \hat{v}, A_1) = (\delta_1 s_1 \hat{v} + \delta_2 v) A_1 + (\delta_1 s_1 z_1(\hat{v}) + \delta_2 z_1(v) - \delta_1 c_1) \times \left[1 - F\left(\frac{\delta_1 c_1}{(\delta_1 s_1 + \delta_2) z_1(\hat{v})}\right)\right]$$

- Savoring $\Rightarrow$ wants to raise $\hat{v}$, but what does it do second term? Two types of situations:

- *Wealth accumulation, status-seeking, entrepreneurial behaviors:* v = ability to accumulate , or to enjoy, material or social assets. Then $z_t(v)\nearrow$ and wishful thinking can help alleviate self-motivation problem: dreams of riches and glory (and of how much will enjoy them) make you work harder.

- *Health, safe driving and other risk-prevention behaviors:* v = immunity from disease, accidents (good genes, driving skills), etc. Then $z_t(v)\searrow$ and wishful thinking leads to "care-free'' complacency / denial that further worsens negligent behavior.

- Date-0 payoffs U : similar to before.

EQUILIBRIUM

• Behavior: probabilities x_L, x_H that someone who gets a signal v that he is either a high-valuation or a low-valuation type invests at $t = 0$.

Optimally chosen, given anticipated costs and benefits, including the hedonic / and or instrumental value of the self-image / identity that will result at $t = 1$.

$$v \in \{v_L, v_H\} \rightarrow \max_{a_0} \left\{ U\left(v, A_0, a_0\right) + \mathbf{V}\left(v, \hat{v}\left(a_0\right), A_0 + a_0 r_0\right) \right\},$$

where $$\mathbf{V}\left(v, \hat{v}, A_1\right) \equiv \lambda V\left(v, v, A_1\right) + \left(1-\lambda\right)\left(v, \hat{v}, A_1\right)$$

brings together "demand" (preferences) and "supply" (cognition) sides of belief formation.

Type $i = H, L$ invests if: $$\mathbf{V}\left(v, \hat{v}(1), A_0 + r_0\right) - \mathbf{V}\left(v, \hat{v}(0), A_0\right) \geq c_0^i$$

• Beliefs: at date 1, when does not have direct recall of his "deep" preferences or values v, (occurs with probability λ), uses own past conduct to try and infer them

In drawing such inferences, individuals are sophisticated (could relax): use Bayes' rule and equilibrium strategies: $\hat{v}\left(a_0\right) = E\left[v | a_0; x_H, x_L\right]$. Perfect Bayesian equilibrium.

• Refinements:

- Monotonicity of beliefs (high type more likely to invest) holds off equilibrium path, as does on it.

- No "self-traps": when multiple equilibria, choose the Pareto dominant one (always exists).

EQUILIBRIUM

• Behavior: probabilities x_L, x_H that someone who gets a signal v that he is high / low valuation type invests at $t = 0$.

Optimally chosen, given anticipated costs and benefits, including the hedonic / and or instrumental value of the self-image / identity that will result at $t = 1$.

$$\mathbf{V}\left(v, \hat{v}(1), A_0 + r_0\right) - \mathbf{V}\left(v, \hat{v}(0), A_0\right) \geq c_0^i$$

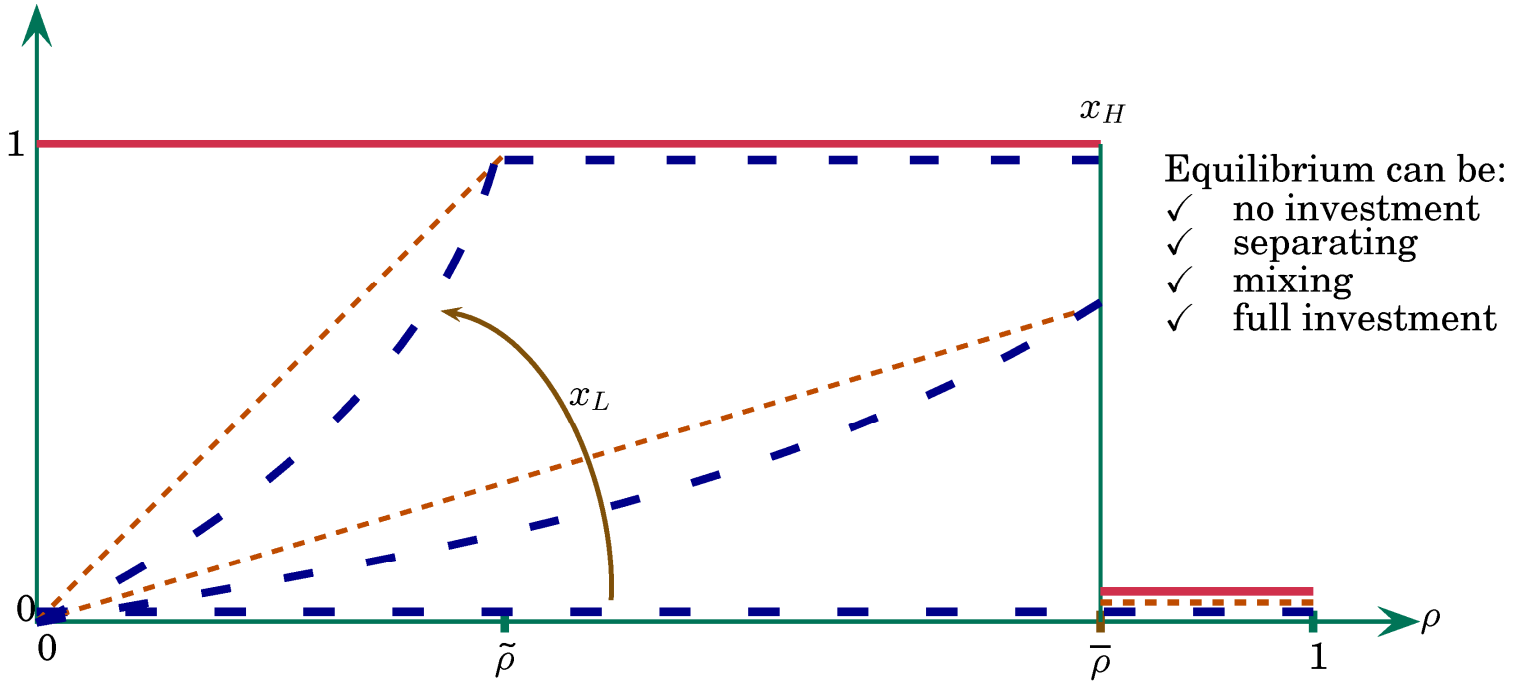
• *Beliefs: in drawing (self-) inferences, individuals are sophisticated / Bayesian.*

EQUILIBRIUM

Proposition 1. *There exists a unique (monotonic, undominated) equilibrium, with thresholds* $\tilde{\rho} \leq \bar{\rho}$ *and investment probabilities* x_L, x_H *such that:*

(1) $x_H(\rho) = 1$ *for* $\rho < \bar{\rho}$ *and* $x_H(\rho) = 0$ *for* $\rho \geq \bar{\rho}$;

(2) $x_L(\rho)$ *is noncreasing on* $[0, \tilde{\rho}]$, *equal to 1 on* $[\tilde{\rho}, \bar{\rho})$ *and equal to 0 on* $[\bar{\rho}, 1]$.



High type's behavior

Low type's behavior

Average behavior

- *Symmetric information bargaining:* at the end of $t = 0$, value of y is revealed to partners, as well as each one's productivity / type v. Decide whether to:
  - stay together ⇒ at $t = 2$, will generate same (expected) y. Bargain now over shares.
  - quit / fight ⇒ each side will get some reservation value v^i, with $v_H > v_L$.

- It is efficient for both "balanced" and "unbalanced" teams to stay together, but in the latter case H partner will require some compensating transfer :

$$y_G > 2v_H > y_B > v_H + v_L > 2v_L$$

- Joint output y is hard data, easy to remember and verify, but *individual contributions* to it – types v – are *soft, unverifiable information* ⇒ later on, *imperfectly recalled* by each side (probability $\lambda < 1$, for simplicity $\lambda = 0$ here).

- Individuals experience *anticipatory feelings* from long-run ($t = 2$) consumption. Same savoring parameter s_I. Could also be pure self-esteem concerns.

  ⇒ incentive to quit / destroy low-productivity match to try and convince oneself that one is an H type / not to blame for the low output.

- Beliefs off the equilibrium path:
  - if only one side requests a share θ_i outside equilibrium set $\Theta \equiv \{\theta_L^*, 1/2, \theta_H\} \Rightarrow$ the other is presumed to have played her equilibrium strategy θ_j^*.
  - if both request the same share $\theta_i = \theta_j \neq 1/2 \Rightarrow$ both get unconditional mean $\bar{v}$
  - if request $\theta_i > \theta_j$ both in $\Theta \Rightarrow$ i gets v_H and j gets v_L (~NWBR)

$\Rightarrow$ *Equilibria*

- High productivity pairs *HH* stay together and split equally. Look at y_L pairs.

- *HL:* unbalanced team: for H partner to accept his share, need

$$(1+s_1)\theta_H^* y_B \geq (1+s_1)v_H, \text{ or } \theta_H^* y_B \geq v_H .$$

For the weak partner to accept his rather than break match, must have

$$(1+s_1)\theta_L^* y_B \geq v_L + s_1\bar{v}$$

$\Rightarrow$ Set of mutually agreeable sharing rules shrinks with s_1 : $\dfrac{v_L + s_1\bar{v}}{1+s_1} \leq \theta_L^* y_B \leq y_B - v_H$

- *LL:* balanced team: viable if: $(1+s_1)y_B/2 \geq v_L + s_1 v_H$

Proposition 8. *All dissolutions are inefficient. Yet, there exist s* and s** such that:*

(1) For $s_1 \leq s^*$, both balanced (LL) and unbalanced (HL) low-output partnerships successfully negotiate, splitting resources equally in the first case and according to any sharing rule in an agreement range that shrinks with s_1 in the second.

(2) For $s^* < s_1 \leq s^{**}$, the two sides can still agree if they share equal blame but not if one must shoulder it all: LL matches survive but HL ones are destroyed.

(3) For $s_1 > s^{**}$, not even balanced (LL) partnerships can find a sustainable agreement.