

## Over My Dead Body: Bargaining and the Price of Dignity

*By ROLAND BÉNABOU AND JEAN TIROLE \**

“If you cut the pay of all but the superperformers, you have a big morale problem. Everyone thinks they are a superperformer”.  
(Head of Human Resources of a manufacturing company, in Truman Bewley (1999)).

“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”.

(Small business owner, in Bewley (1999)).

Concerns of pride, dignity and the desire to “keep hope” about future options often lead individuals and groups to walk away from reasonable offers, try to shift blame for failure onto others, or take refuge in political utopias, leading to impasses and conflicts. Examples include trials, divorces, strikes, the scapegoating of minorities for economic hardships, and war. A key and puzzling aspect of these processes is the role played by wishful *rationalizations and delusions*, as attested by field observers (e.g., Bewley (1999) in the context of labor relations, Woods, Lacey and Murray (2006) in that of war) as well as controlled experiments. Thompson and Loewenstein (1992) and Babcock et al. (1995) thus demonstrate how subjects in bargaining situations with common knowledge spontaneously generate, through self-serving processing and recall of the same evidence, divergent beliefs about the fairness of their cause and wishful predictions of outcomes, and how these are associated to costly delays and disagreements.

\* Bénabou: Department of Economics and Woodrow Wilson School, Princeton University, Princeton, NJ 08544, CIFAR, CEPR, IZA and NBER. Tirole: Toulouse School of Economics, 21 Allées de Brienne, 31000 Toulouse, France, and MIT. We are grateful for helpful comments to Andrew Caplin, Robert Oxbury, Philipp Sadowski and Glen Weyl, as well as to participants at several conferences and seminars. Bénabou gratefully acknowledges support from the Canadian Institute for Advanced Research.

To analyze these behaviors, we propose here a simple model of how anticipatory or self-esteem concerns lead to the inefficient breakdown of Coasian bargaining under *symmetric information*, as both sides seek to self-enhance by turning down “insultingly low” offers. To do so, we build on Roland Bénabou and Jean Tirole (2007), which develops a general framework for analyzing social and economic phenomena involving beliefs which people “invest in”.

The underlying idea is that individuals are often uncertain or insecure about their own “deep values”, abilities or worth and that, having better, more objective access to the track record of their *actions* than to the exact mix of *motivations* that spurred them, they are rationally led to judge themselves by what they do.<sup>1</sup> When contemplating choices, they then factor in what kind of a person each alternative would “make them” and the desirability of those self-views. The theory is thus cognitive, as it explicitly models identity and related concepts as beliefs and emphasizes the *self-inference* process through which they operate. At the same time, the *value* of identity or dignity arises because they confer affective benefits, functional ones, or both. The first case arises when self-esteem has pure consumption value or when future prospects give rise to anticipatory feelings such as savoring or dread. The second obtains when a strong sense of self provides clear priorities and directions that help the individual mobilize energy and resist short-term temptations.

From these two basic assumptions –self-inference and motivated beliefs– a broad range of results follow, even in single-agent or non-strategic contexts. Identity investments increase when information is scarce (e.g., new situations) or when a greater relative en-

<sup>1</sup>See, e.g., Festinger and Carlsmith (1959) on cognitive dissonance, Bem (1972) on self-perception and Quattrone and Tversky (1984) on the self-manipulation of “diagnostic” actions. For recent experiments on the strategic management of self-image through costly actions or information-avoidance, see Dana, Weber and Kuang (2003) and Mazar, Amir and Ariely (2006).

dowment of some asset (wealth, career, family, culture) raises the stakes on viewing it as valuable. Taboos against certain transactions or even the mere contemplation of tradeoffs arise to protect fragile beliefs about the “priceless” value of certain assets (life, freedom, love, faith) or things one “would never do”. Escalating commitments can lead to a “hedonic treadmill”, and competing identities cause dysfunctional failures to invest in high-return activities (education, assimilation) or even the destruction of productive assets. In social interactions, norm violations elicit a forceful response (exclusion, harassment) when they threaten a strongly held identity, but further erode morale when it was initially weak.

We extend here this framework to bargaining and other distributive conflicts. We consider a partnership of two individuals or groups (parties in a legal dispute, capital and labor, majority and minority populations) who must decide whether to continue together or destroy the match. Continuation always yields a positive surplus, but a low output realization means that at least one party has low ability. Moreover, whereas joint output is hard data, individual contributions to it (“who is to blame”, “who is getting a raw deal”) are soft signals, symmetrically observed when bargaining but imperfectly recalled following a split. Agreeing to inferior or even equal contractual terms in a low-performance team then entails a loss in self image and / or anticipatory utility. Conversely, by refusing “insultingly low” offers and destroying the match when they do not obtain enough of a concession, each side can try to preserve or salvage their dignity and shift the blame onto the other, taking refuge from bleak realities in feelings of self-righteousness and wishful hopes for “a better tomorrow”. In equilibrium, the range of sustainable sharing rules is shown to shrink with the importance of self-image or anticipatory concerns. Beyond a point, a bargaining impasse becomes unavoidable, in spite of gains from trade and fully symmetric information.

The paper relates to three literatures. The first one concerns cognitive dissonance and motivated beliefs (e.g., Akerlof and Dickens (1982), Rabin (1995), Carrillo and Mariotti (2000), Bénabou and Tirole (2002, 2006a), Köszegi (2004), Battaglini, Bénabou and Tirole (2005), Brunnermeier and Parker (2005)), as well as the related issue of anticipatory feelings (e.g., Loewenstein (1987), Caplin and Leahy (2001)). Most closely related, through the idea of self-signaling or self-reputation, are Bodner and Prelec (2003), Bénabou and Tirole (2004, 2006b), Bernheim

and Thomadsen (2005), Young (2006) and Dal Bó and Terviö (2008). On the experimental side, Konow (2000) and Dana, Kuang, and Weber (2003) demonstrate that individuals making monetary allocations affecting their own payoffs engage in self-deception and information avoidance about the fairness or likelihood of other players’ outcomes. To our knowledge, none of this literature has directly addressed bargaining situations.

The second body of work is that on identity, particularly Akerlof and Kranton (2000, 2005) and Oxoby (2003, 2004). In these models, agent’s preferences (or perceptions represented as utility parameters not directly tied to an information structure) depend on their endogenously chosen group memberships. We instead explicitly model the management of beliefs and the cognitive mechanisms through which it occurs. This also leads to different results, such as the fact that being able to manage his own identity can often make a person worse off.

Finally, there is also a recent literature on bargaining and contracting with heterogeneous beliefs (e.g., Yıldız (2003), (2004), Ali (2006)). The general motivation of these papers is also to understand the sources of delays and breakdowns, but its methods and focus are quite different. In particular, beliefs are exogenous (e.g., arising from given priors) and remain invariant to offers and counteroffers; they also bear on the bargaining process (stochastic recognition) rather than on parties’ outside options. On the other hand, these papers make explicit the dynamic aspects of bargaining, whereas we consider a much simpler Nash demand game.

## I. Model

### A. Technology

We consider a “partnership” between two risk-neutral individuals or groups –spouses, labor and management, majority and minority populations, etc. Each may be of high or low type, corresponding to ability, motivation, honesty, deservedness, outside opportunities, etc: for  $i \in \{1, 2\}$ ,

$$(1) \quad v^i = \begin{cases} v_H & \text{with probability } \rho \\ v_L & \text{with probability } 1 - \rho \end{cases},$$

with  $v_H > v_L$ . There are three periods, as illustrated in Figure 1, and we abstract from discounting. At date 0, the joint output or productivity of

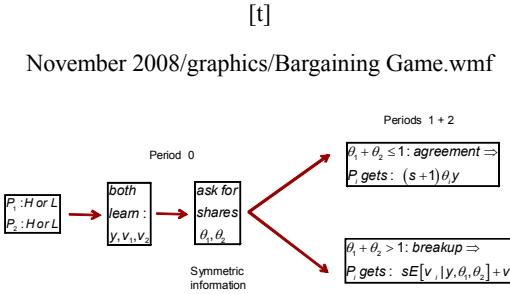


FIGURE 1: BARGAINING WITH MALLEABLE BELIEFS

the partnership is revealed: it is either good or bad,  $y \in \{y_B, y_G\}$ , with  $y_G > y_B$ . The technology exhibits complementarity, in that  $y = y_G$  if and only if  $v^1 = v^2 = v_H$ . The interesting case will then be when  $y = y_L$ , since this means that at least one of the parties is “to blame” for the low output – disappointing marriage, firm or economy, lost war, etc.

At the end of period 0, the two partners must decide whether to: (i) remain together, in which case they will continue to produce the same (expected) output in period 2 (the long run) and must bargain over how it will be shared; or (ii) split, in which case each side will get a reservation value that, for simplicity, corresponds to his type  $v^i$ : producing in autarky, searching for a new match, or triggering a costly fight with the other side for control of resources.

Let parameters be such that staying together is efficient for all teams, both balanced (*HH* or *LL*) and unbalanced (*HL*), but in the latter case a compensating transfer (or share of  $y_B$  exceeding 1/2) is needed to induce the more productive partner to stay:

$$(2) \quad y_G > 2v_H > y_B > v_H + v_L > 2v_L.$$

When bargaining and making their stay or quit decisions at the end of period 0, the two parties are assumed to know not only the joint output *y*, but also each one’s type. Such *common knowledge* will make inefficient-breakdown results all the more interesting and allow us to provide a formal model of the Babcock et al. (1995) types of findings described above.

### B. Preferences and beliefs

In keeping with our general self-inference approach to identity, we further assume that, at date 1 :

(i) Whereas the level of joint output *y* is “hard” data that is easy to remember and verify, individuals’ separate contributions to it –their types *v*– represent soft, unverifiable information, which later on is only imperfectly recalled.<sup>2</sup> Indeed, it would always be more pleasant, *ceteris paribus*, to “recall” that one was the competent and honest partner and the other was entirely to blame for the team’s poor performance (“everyone thinks they are a superperformer”).

(ii) Individuals experience anticipatory feelings, such as hope and dread, from their long-run (date-2) income or consumption prospects. Alternatively, they may derive utility from pure self-esteem about their talent or worth; this slight variant leads to very similar results.

We now formalize and discuss further each of these two premises.

For a person’s past choices to define his sense of identity or dignity, it must be that these are *informative* about the “kind of person” he is, and therefore that he himself is, at times, not fully confident of his own type –deep values, abilities, etc. Similarly, if he later perfectly understood that what tipped the scales on a decision was the desire to achieve a certain self-image, such attempts would come to nil. Some form of imperfect self-knowledge (memory, accessibility) is therefore essential to understanding how people’s choices can be shaped by concerns such as “being true to myself,” “maintaining my integrity,” “keeping my self-respect”, etc. And indeed, there is extensive evidence that people’s recall of their past feelings, efforts and motivations is highly imperfect and self-serving, that they judge themselves by their behavior, and consequently tailor the latter to preserve certain self-views.<sup>3</sup>

**ASSUMPTION 1: (Self-inference).** At date 1, each player is aware (or reminded) of past individual contributions,  $v^i$ ,  $i = 2$ , only with probability  $\lambda$ . With probability  $1 - \lambda$ , he no longer recalls (has access

<sup>2</sup>Given the same information, subjects in bargaining situations systematically recall more of the evidence that favors their own side, even when roles are exogenously determined (Thompson and Loewenstein (1992)). In dictator games, they take advantage of contextual ambiguity to “persuade” themselves that they deserve more than what they judge to be the fair share when making allocations between other people (Konow (2000)).

<sup>3</sup>See again footnote 1. Further discussions and references can be found in Bodner and Prelec (2003), Bénabou and Tirole (2004, 2007) and Battaglini et al. (2005).

to) these signals and uses instead the outcome of the negotiation to infer his and the other player's types.

We shall denote by  $\hat{\rho}^i$  individual  $i$ 's date-1 belief about "what kind of a person" he is and by  $\hat{v}^i \equiv \hat{\rho}^i v_H + (1 - \hat{\rho}^i)v_L$  the corresponding expected type, either of which defines his (subjective) sense of identity. With probability  $\lambda$ , the posterior  $\hat{v}^i$  is thus equal to the true type (or unbiased signal)  $v^i$ , and with probability  $1 - \lambda$  it is equal to the conditional expectation  $\hat{v}^i \in [v_L, v_H]$  that can be inferred from what offers were made and whether they were accepted or rejected. We assume that, in making these inferences at  $t = 1$ , players are fully rational Bayesians. Although this assumption can easily be relaxed, it is a natural benchmark and imposes discipline on the extent to which agents can choose to believe what suits them.<sup>4</sup>

What suits them, in turn, depends on the affective needs or instrumental functions goals that identity or dignity serves for them. As discussed in Bénabou and Tirole (2002, 2007), the former include pure ego-gratification as well as remaining hopeful about one's future prospects (anticipatory utility); the latter include the motivational value of "believing in oneself" to achieve long-term goals and overcome self-control problems, as well as a possible facilitating role in signaling to others (if it is easier to persuade others of a claim, true or false, when one is convinced of it). We shall focus here on the first class of motives, namely "mental consumptions" (Schelling (1985)), but also explain in Section E how a simple variant yields a functional role for dignity, which strengthens the will to resist temptations at date 1.

In what follows, we denote by  $E_t^i$  an agent  $i$ 's expectations at date  $t = 0, 1$ .

<sup>4</sup>It also makes the model directly applicable to contexts where the two bargaining parties are signaling to an outside audience. Such social-reputational concerns, however, are "shut off" (in particular, through anonymity) in all the experimental evidence (e.g., Bem (1972), Quattrone and Tversky (1984), Thompson and Loewenstein (1992), Babcock et al. (1995), Konow (2000), Dana et al. (2003), Mazar et al. (2006)). In many field surveys, they also seem secondary in importance to individuals' self-perceptions (see, e.g., the above quotations from Bewley (1998)). Thus, although self-reputation and social reputation are very complementary concerns, they correspond to empirically distinct phenomena and their analyses point to different mediating mechanisms –in particular, the key role of memory / retrospective accessibility in the pursuit of self-serving beliefs.

ASSUMPTION 2: (*Motivated beliefs*). Let  $U_2^i$  denote agent  $i$ 's long-run income, equal to  $\theta^i y$  when bargaining leads to agreement and to  $v^i$  when it leads to a split. At  $t = 0$ , each agent seeks to maximize the (undiscounted) expected present value

$$(3) \quad U_0^i \equiv E_0^i [su_1^i + U_2^i],$$

where  $u_1^i$  is a utility flow received during period 1 and equal to either:

- (i)  $u_1^i = E_1^i [U_2^i]$ , in the anticipatory-utility case
- (ii)  $u_1^i = E_1^i [v^i]$ , in the pure self-esteem case.

As made clear by our notation, the two cases are closely related. We shall focus the exposition on (i), which is somewhat more "consequentialist", but all the paper's results are qualitatively identical with (ii).

### C. Bargaining

We formalize the bargaining process as a standard Nash demand game. At  $t = 0$ , with full and symmetric information, players 1 and 2 simultaneously make demands for shares  $\theta_1$  and  $\theta_2$  of future output,  $y$ .<sup>5</sup> A larger share may correspond to a monetary transfer, a control right (regional autonomy, child custody, seats on the board) or a new performance measurement system that will alter the sensitivity of income shares to individual contributions. If  $\theta_1 + \theta_2 \leq 1$  each gets what they asked for, whereas if  $\theta_1 + \theta_2 > 1$  the negotiation breaks down and the pair dissolves. We assume that offers are later remembered (having been formally recorded, submitted to an arbitrator, etc.), but the key results are similar when they are not.

We look for a symmetric, pure-strategy Perfect Bayesian equilibrium, with shares  $\theta_H^* > 1/2 > \theta_L^*$  for the high and low types respectively in an unbalanced partnership, and a common share  $1/2$  in a balanced one.<sup>6</sup> Finally, we restrict out-of-equilibrium beliefs as follows. Let  $\Theta \equiv \{\theta_L^*, 1/2, \theta_H^*\}$  denote the set of equilibrium offers. For  $\theta_i \in \Theta$  and  $\theta_j \notin \Theta$ ,

<sup>5</sup>We treat the allocation of period-0 output (if any) as sunk –e.g., shared *ex ante* on a 50-50 basis, before types are revealed. Since expected output is equal in both periods, allowing initial resources to be part of the bargaining would simply amount to doubling the size of the pie.

<sup>6</sup>Due to imperfect recall, this is formally a game with four players: each agent's "self 0" is engaged in a signalling game with each other his "self 1".

player  $i$  is presumed to have played on the equilibrium path; together with the value of  $y$ , this is sufficient to tie down both players' types. If  $\theta_i$  and  $\theta_j$  are both in  $\Theta$  but are jointly inconsistent with equilibrium, on the other hand, then: (i) if  $\theta_i = \theta_j$  (e.g., both sides demand  $\theta_H^*$ ) the two players are considered equally likely to have deviated, resulting in  $\hat{v}_i = \hat{v}_j = (v_H + v_L)/2 \equiv \bar{v}$ ; (ii) if  $\theta_i > \theta_j$ , then  $\hat{v}_i = v_H$  and  $\hat{v}_j = v_L$ ; this is in the spirit of standard equilibrium refinements (such as D1), since it is always the strong type who has less to lose from breaking up the match.

#### D. Equilibrium

Let us first observe that in any equilibrium with agreement, the shares demanded by both sides must sum to 1. Otherwise, either party can ask for  $\epsilon$  percent more and gain  $(1+s)\epsilon y$ , since the team will still stay together. For the same reason, downward deviations by either type (asking for less than the equilibrium share) are never profitable. The binding constraints will thus correspond to upward deviations.

Since  $(1+s)y_G/2 > (1+s)v_H > (v_H + s\hat{v})$  for any feasible value of  $\hat{v}$ , matched strong partners ( $HH$ ) always stay together, sharing output equally. The interesting case is that of low-productivity pairs,  $y = y_B$ . Consider first bargaining in an unbalanced ( $HL$ ) team. For the  $H$  type to be satisfied with his share, it must be that:

$$(4) \quad \theta_H^* y_B \geq v_H.$$

Otherwise he could ask for more, which would break up the team while maintaining his posterior belief  $\hat{v} = v_H$  (since the other's offer of  $\theta_L^* < 1/2$  is recalled) and achieving  $(1+s)v_H > (1+s)\theta_H^* y_B$ .

Next, for the weak partner ( $L$  type) to accept the bargain, it must be that:

$$(5) \quad (1+s)\theta_L^* y_B \geq v_L(1+\lambda s) + s(1-\lambda)\bar{v},$$

otherwise he could deviate and break the match by demanding  $\theta_H^*$  (mimicking the strong partner), thus achieving with probability  $1 - \lambda$  the posterior self-view  $\hat{v} = \bar{v}$ , even though his true "worth" and outside option is only  $v_L$ . Other deviations to  $\theta' > \theta_L$  with  $\theta' \neq \theta_H^*$  would still identify him as the weak type,  $\hat{v} = v_L$ , and be *a fortiori* unprofitable under (5). The set of mutually agreeable sharing rules  $(\theta_L^*, 1 - \theta_L^*)$  is

[t]

November 2008/graphics/Agreement Set.wmf

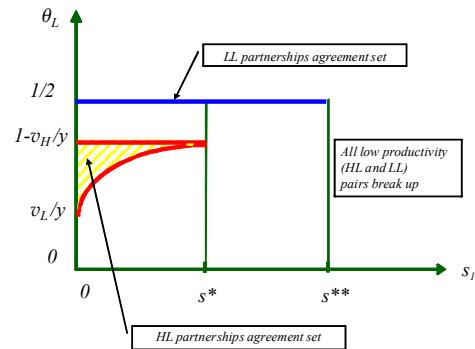


FIGURE 2: BARGAINING SETS AND BREAKDOWN REGIONS (FOR  $s^* < s^{**}$ )

thus defined by

$$(6) \quad \frac{v_L(1+s\lambda) + s(1-\lambda)\bar{v}}{1+s} \leq \theta_L^* y_B \leq y_B - v_H.$$

As illustrated in Figure 2, it shrinks as identity concerns increase, up to

$$(7) \quad s^* \equiv \frac{y_B - v_H - v_L}{v_H + \lambda v_L + (1-\lambda)\bar{v} - y_B}$$

when  $y_B < v_H + \bar{v}$  (otherwise, let  $s^* \equiv +\infty$ ). Beyond this critical threshold a *bargaining impasse arises*, in spite of gains from trade and symmetric information. Intuitively, a higher  $s$  makes the loss of self-image involved in "admitting blame" more costly for the  $L$  type, who then requires a higher share  $\theta_L^*$  to be compensated. At some point this becomes more than the  $H$  type is willing to grant given his outside option, and no agreement can be reached. The two parties then split (or fight) by both demanding  $\theta_H^*$  (which is now a dominant strategy).

We next turn to bargaining in an  $LL$  team. By asking for a share  $\theta' > 1/2$ , either side can break up the match and achieve, with probability  $1 - \lambda$ , a self-image  $v_H$ . Therefore, the partnership remains sustainable only if  $(1+s)y_B/2 \geq v_L + s[\lambda v_L + (1-\lambda)v_H]$  or

$s \leq s^{**}$ , where

$$(8) \quad s^{**} = \frac{y_B - 2v_L}{2[\lambda v_L + (1-\lambda)v_H] - y_B}.$$

Otherwise the match is dissolved, as each side seeks to convince himself that he is better than the other (demanding again  $\theta_H^*$ ), even though in reality both are equally bad; see again Figure 2.

Depending on parameter values,  $s^{**}$  can be above  $s^*$ , as illustrated in Figure 2, or below it.<sup>7</sup> For brevity, we shall focus on the former. Straightforward but tedious derivations show that  $s^* < s^{**}$  if and only if  $3y_B/2 < 2v_H + v_L$ . Together with (2), this means that  $v_H + v_L < y_B < (2/3)(v_H + v_L)$ .

We can obtain a further result by naturally linking joint output to individual productivities. Consistent with our earlier assumptions, let  $HL$  and  $LL$  pairs both produce  $y_B = \Phi v_L$ , where  $\Phi$  is such that (2) holds. It is then simple to verify that, as  $v_H/v_L$  rises,  $s^*$  and  $s^{**}$  both decrease and (6) becomes more stringent.

**PROPOSITION 1:** (1) For  $s \leq s^*$ , unbalanced low-output ( $HL$ ) partnerships successfully negotiate, splitting resources according to any sharing rule  $\theta_L^*$  satisfying (6). This agreement range shrinks with  $s$  and, for  $s > s^*$ , the match is inefficiently destroyed. (2) For  $s \leq s^{**}$ , balanced low-output ( $LL$ ) partnerships successfully negotiate, splitting resources equally. For  $s > s^{**}$ , the match is inefficiently destroyed. (3) Let  $y_B = \Phi v_L$ . For any  $s$ , the bargaining set shrinks and both types of impasses become more likely, the greater the inequality  $v_H/v_L$  between high and low types' productivities.

Our model of bargaining with malleable beliefs identifies a new and potentially important limit to the achievement of Coasian deals, namely the preservation of dignity, pride, or "hope" about the future. It also leads to testable predictions, as both salience  $s$  and the productivity differential  $v_H/v_L$  can be ma-

<sup>7</sup>Intuitively, the high types lose less in material payoffs from breakup than low types, which tends to make  $HL$  pairs less stable than  $LL$  ones. On the other hand, the reputational gain which a low type can achieve from breakup is higher in an  $LL$  pair (by refusing the other's side "moderate" offer of 1/2 and demanding  $\theta' > 1/2$  instead, he can "pass" as a high type) than in an  $HL$  one (where he can only attain  $\hat{v} = \bar{v}$ , by asking for the same  $\theta_H^*$  as his partner).

nipulated experimentally. The latter can also be measured empirically in real-world contexts, where one should observe that more unequal bargaining positions reduce the likelihood of agreement.

From (7) and (8), we also have:

**PROPOSITION 2:** *A breakdown of Coasian bargaining is more likely:*

- (i) *The more salient are agents' identity concerns (higher  $s$ ).*
- (ii) *The more malleable are their memories, and hence their beliefs (the lower  $\lambda$ ).*

### E. Welfare

By our first equilibrium refinement, when  $HL$  pairs split both sides must be asking for the same  $\theta_H^* > 1/2$ , and when  $LL$  pairs also split the same must hold. Otherwise, one side can deviate to  $\theta_H^*$  and achieve the maximal self-reputation  $v_H$ . In any pair that splits, therefore, each side ends up with  $v^l(1+s\lambda) + s(1-\lambda)\bar{v}$ , where

$$(9) \quad \tilde{v} \equiv E[v | y_B, \theta^1 = \theta^2 = \theta_H^*]$$

is the average value of  $v$  over all such dissolutions. There is thus, *in fine*, no net gain in self-esteem or anticipatory utility, only a transfer from the high to the low type within  $HL$  pairs, and from  $HL$  to  $LL$  pairs when the latter also break up. Indeed the management of self-image is a *zero-sum game*, by the martingale property of (Bayesian) beliefs. What remains, however, is the *destruction of surplus* brought about by the quest for such gains.

To identify the latter, consider first the region  $s^* < s < s^{**}$ , in which only  $HL$  pairs split. Then  $\tilde{v} = (v_H + v_L)/2 = \bar{v}$ , and the average loss over all  $\rho(1-\rho)$  such pairs is

$$(10) \quad \begin{aligned} &y_B(1+s) - (v_H + v_L)(1+s\lambda) \\ &- 2(1-\lambda)s\bar{v} \\ &= (1+s)(y_B - 2\bar{v}) > 0. \end{aligned}$$

For  $s > s^{**}$  both  $HL$  and  $LL$  pairs split, so  $\tilde{v} = (\rho/2)v_H + (1-\rho/2)v_L$  and the average loss over all  $1-\rho^2$  such dissolutions is

$$(11) \quad \begin{aligned} &y_B(1+s) - 2s(1-\lambda)\tilde{v} \\ &- (1+s\lambda)[\rho(v_H + v_L) + (1-\rho)2v_L] \\ &= (1+s)[y_B - 2\tilde{v}] > 0. \end{aligned}$$

**PROPOSITION 3:** *An increase in the malleability of beliefs  $1-\lambda$  always reduces ex-ante welfare. The same holds for an increase in the salience  $s$  of anticipatory utility or identity concerns.*

In Bénabou and Tirole (2007) we show that, whereas the positive implications of individual belief management are very similar whether it arises from hedonic motives (self-esteem, anticipatory feelings) or instrumental ones (sense of direction, self-discipline), the welfare implications, by contrast, depend critically on this distinction. A similar principle applies in the present *strategic* context. Due to space constraints, we only sketch here this variant of the bargaining model that leads to a more attractive role (normatively speaking) for dignity concerns.

The only additional assumption is that, at date 1, each individual may need to carry out a task that:

- (i) requires costly effort or perseverance, but is potentially subject to a self-control problem (e.g., due to hyperbolic discounting,  $\beta < 1$ );
- (ii) has an expected return that increases with the agent's individual productivity  $v$ , so that perseverance and self-view  $\hat{v}$  are complements.

The date-1 task may be independent of whether the agent is paired or unpaired at that time, or it could apply only to unpaired agents: searching for better opportunities, fighting, or holding out longer in costly bargaining.

In such settings, the enhancement and preservation of dignity by rejecting "realistic" bargaining offers can have positive efficiency effects at date 1. Such pooling at  $t = 0$  boosts the  $v_L$  type's self-confidence and subsequent motivation, although it simultaneously weakens that of the  $v_H$  type. The first effect leads to a welfare gain, the second to a loss. Therefore when the nature of the date-1 self-control problem (value or probability distribution of  $\beta$ ) makes it more of a concern for the low type than for the high one, meaning that its severity is moderate, there is a *net social gain* from the malleability of beliefs ( $\lambda < 1$ ) and the enhancement of the low types' dignity that it allows. When the self-control problem is harder, however, meaning that it affects the high types more often than the low ones, there is again a net social loss.

More generally, for dignity investments to be socially beneficial, it must be that full information about each agent's type is suboptimal from the point of view of date 1. This can arise at the individual level if agents' self-esteem  $\hat{v}$  enters their utility in a nonlinear,

concave manner; or, more interestingly, from a public-goods problem, provided that: (i) individual contributions and self-worth are complements; (ii) underprovision of effort by low types is, on average, more of a concern than it is from the high types.

## II. Conclusion

A simple model was proposed to analyze the role, in bargaining and other distributive conflicts, of *endogenously arising belief distortions* linked to pride, dignity or wishful thinking about future outcomes. A first set of further applications may include contracts and organizational design. For instance, the event in which a conflict arises between material efficiency and dignity preservation (i.e., (2) holds) could be a particular state of the world that only both parties can verify. There is then no (enforceable) contract they can sign that will prevent opportunistic behavior and a breakdown of the match when this contingency arises (Hart and Moore (2008)). A second interesting direction is the political economy of reforms such as opening to trade or liberalizing the labor market. Whereas the standard concern is whether winners can credibly commit to compensating losers, a potentially equally important one is that the latter precisely *do not want to see themselves* (and be identified by others) as losers, now dependent on "handouts" from the rest of nation.

## REFERENCES

- Ali, Nageeb (2006) "Waiting to Settle: Multilateral Bargaining with Subjective Biases," *Journal of Economic Theory*, 130(1), 109-137.
- Akerlof, George A. and Dickens, William T. "The Economic Consequences of Cognitive Dissonance." *American Economic Review*, June 1982, 72(3), 307-319.
- and Kranton, Rachel E. "Economics and Identity." *Quarterly Journal of Economics*, August 2000, 115(3), 716-753.
- (2005) "Identity and the Economics of Organizations," *Journal of Economic Perspectives*, 19, 9-32.
- Babcock, L., Loewenstein, G., Issacharoff, S. and Camerer, C. (1995) "Biased Judgments of Fairness in Bargaining," *American Economic Review*, 85(1), 1337-1343.
- Battaglini, M., Bénabou, R., and J. Tirole (2005) "Self-Control in Peer Groups," *Journal of Economic Theory*, 123: 105-134.
- Bem, D. J. (1972). "Self-Perception Theory," in

- L. Berkowitz, ed., *Advances in Experimental Social Psychology*, vol. 6, 1-62. New York: Academic Press.
- Bénabou, R. and J. Tirole (2002) "Self Confidence and Personal Motivation," *Quarterly Journal of Economics*, 117(3): 871-915.
- (2004) "Willpower and Personal Rules," *Journal of Political Economy*, 112(4): 848-886.
- (2006a) "Belief in a Just World and Redistributive Politics," *Quarterly Journal of Economics*, 121(2), 699-746.
- (2006b) "Incentives and Prosocial Behavior," *American Economic Review*, 96(5), 1652-1678. [?]
- (2007) "Identity, Dignity and Taboos: Beliefs as Assets," CEPR Discussion Paper 6123, January 2007.
- Bernheim, D. and R. Thomadsen (2005) "Memory and Anticipation," *The Economic Journal*, 115, 271-304.
- Bewley, T. (1999) *Why Wages Don't Fall During a Recession*. Harvard University Press.
- Bodner, R. and D. Prelec (2003) "Self-signaling and Diagnostic Utility in Everyday Decision Making," in I. Brocas and J. Carrillo eds. *The Psychology of Economic Decisions. vol. 1: Rationality and Well-being*. Oxford University Press.
- Brunnermeier, M. and J. Parker (2005) "Optimal Expectations," *American Economic Review*, 95, 1092-1118.
- Caplin, A., and J. Leahy (2001) "Psychological Expected Utility Theory and Anticipatory Feelings," *Quarterly Journal of Economics*, 116, 55-80
- Dal Bó, Ernesto and Marko Terviö (2008) "Self-Esteem, Moral Capital and Wrongdoing," NBER Working Paper 14508, November.
- Dana, J., Kuang, J., and R. Weber (2003) "Exploiting Moral Wriggle Room: Behavior Inconsistent with a Preference for Fair Outcomes." Carnegie Mellon Behavioral Decision Research Working Paper No. 349, June.
- Hart, Oliver and Moore, John (2008) "Contracts as Reference Points", *Quarterly Journal of Economics*, 123(1), February, 1-48.
- Hart, Oliver and Moore, John "Incomplete Contracts and Ownership: Some New Thoughts", *American Economic Review*, 97(2), May (2007), 182-86. [?]
- Kay, A., Jimenez, M. and J. Jost (2002) "Sour Grapes, Sweet Lemons, and the Anticipatory Rationalization of the Status Quo," *Personality and Social Psychology Bulletin*, 9, 1300-1312.
- Konow, J. (2000) "Fair Shares: Accountability and Cognitive Dissonance in Allocation Decisions," *American Economic Review*, 90(4), 1072-1091.
- Köszegi, B. (2004) "Utility from Anticipation and Personal Equilibrium," U.C. Berkeley mimeo, June.
- Kunda Z. (1987) "Motivated Inference: Self-Serving Generation and Evaluation of Causal Theories," *Journal of Personality and Social Psychology*, 53(4), 636-647.
- Lamont, M. (2004) *The Dignity of Working Men: Morality and the Boundaries of Race, Class and Immigration*. Harvard University Press.
- Loewenstein, G. (1987) "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 97, 666-84.
- Mazar, N., Amir, O. and D. Ariely (2006) "Mostly Honest: A Theory of Self-Concept Maintenance," MIT mimeo, December.
- Oxoby, R. (2003) "Attitudes and Allocations: Status, Cognitive Dissonance and the Manipulation of Preferences," *Journal of Economic Behavior and Organization*, 52(3), 365-385.
- (2004) "Status, Cognitive Dissonance, and the Growth of the Underclass," *The Economic Journal*, 114(498), 727-749.
- Quattrone, G., and Tversky, A. (1984) "Causal Versus Diagnostic Contingencies: On Self-Deception and the Voter's Illusion," *Journal of Personality and Social Psychology*, 46(2), 237-248.
- Rabin, M. (1995) "Moral Preferences, Moral Constraints, and Self-Serving Biases," Berkeley Department of Economics Working Paper No. 95-241, August.
- Schelling, T. (1985) "The Mind as a Consuming Organ." In J. Elster (Ed.), *The Multiple Self*. New York: Cambridge University Press, 177-195.
- Thompson, L. and G. Loewenstein (1992) "Ego-centric Interpretations of Fairness in Negotiation," *Organization Behavior and Human Decision Processes*, 51, 176-197.
- Woods, K., Lacey, J. and W. Murray (2006) "Saddam's Delusions: The View from the Inside," *Foreign Affairs*, June
- Yildiz, M. (2003) "Bargaining Without a Common Prior – An Immediate Agreement Theorem," *Econometrica*, 71 (3), 793-811.
- Yildiz, M. (2004) "Waiting to Persuade," *Quarterly Journal of Economics*, 119 (1), 223-249.