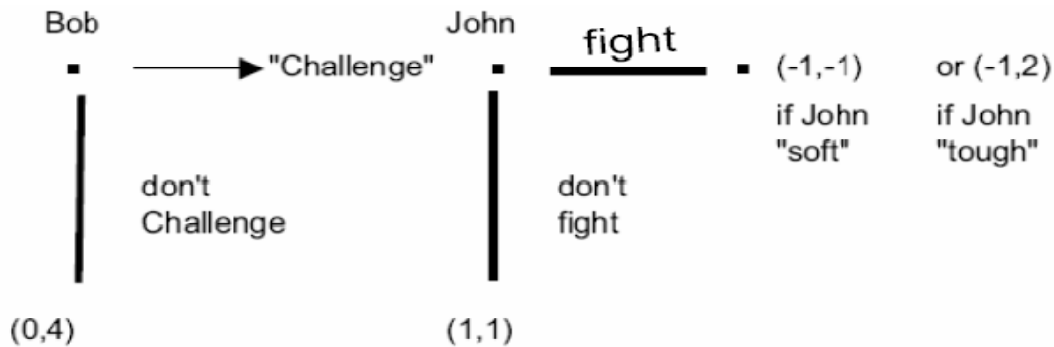


Lecture Note on Campaigning for Respect (and Being Judged for It)
Prof. Glenn C. Loury, EC 1370, February 25, 2010

The Reputation Game models an ethically significant the interaction between Bob and John. (See diagram below.) In this dynamic game of incomplete information, John can be either of two types – soft or tough (one thinks of Elijah Anderson’s “decent” vs. “street” distinction.) Bob chooses whether to Attack or Not; John responds, choosing whether or not to Fight. (The game is ‘ethically significant’ because it shows that the link between character and behavior depends on the social context.) The payoffs (specified in parentheses in the diagram) are such that both the soft and the tough versions of John would much prefer not to be attacked, though if attacked a soft John would want not to fight and a tough John would want to fight. (This is reflected in the fact that both soft and tough versions of John receive payoff = +1 from not fighting, while fighting gives soft John a payoff = -1 but tough John’s payoff from fighting =+2.) Also, the payoffs indicate that Bob wants not to attack if will be fought (payoff = -1 versus 0), and wants to attack if he will not be fought (payoff = +1 versus 0).

THE REPUTATION GAME



Payoffs =(Bob, John)

If this interaction between Bob and John were to occur only once, then rational agents would play as follows: if attacked, a tough John fights and a soft John does not. So, Bob computes the average of his payoffs over the outcomes that could obtain, and attacks only if he thinks the probability that John is tough is less than 1/2. The soft and tough types of John react naturally. The game is trivial.

But, should these players interact twice in succession, the outcome of the game is more interesting. Now, John’s action at the first stage can serve as a signal to Bob about his type, thereby affecting how the second stage is played. John knows this. Bob knows that John knows, John knows that Bob knows that he knows..., etc.

Given this structure, we make the following two claims:

Claim 1: In the twice-played game it is inconsistent with rationality for the soft John never to fight when first attacked.

Proof: If soft John were never to fight when first attacked, then because tough John always fights, John’s first stage action would be a perfect signal of his type: “fight at the first stage” would mean he’s tough, “not fight” would mean he’s soft. But then,

Bob's rational response to this situation is to attack at the second stage only if he is fought at the first stage. Yet, this response on Bob's part means that soft John could avoid being attacked at the second stage by fighting at the first stage which, if he is rational, he would want to do since $(-1+4)$ —his payoff from fighting and thereby avoiding subsequent attack, exceeds $(1+1)$ —his payoff from not fighting at either stage. This contradicts the supposition that he never fights when first attacked.

Claim 2: In the twice-played game it is inconsistent with rationality for Bob to always attack at the second stage after being fought at the first stage.

Proof: If Bob always attacks after being fought, soft John can gain nothing by fighting at the first stage and so, being rational, soft John would never fight when first attacked. But, Claim 1 asserts that this can't occur when the players are rational.

Taken together, Claims 1 and 2 imply that the only outcome of the twice-played game consistent with player rationality has soft John mixing his behavior at the first stage between fighting and not fighting, while Bob mixes at the second stage between attacking and not, given that he has been initially fought. (If Bob is not fought at the first stage then he knows John is soft and so definitely attacks at the second stage.) Indeed, if $P < 1/2$ is Bob's assessment of likelihood that John is tough, then the unique equilibrium of the twice-played game entails soft John fighting when first attacked with probability $P/(1-P)$, and Bob attacking at the second stage if fought at the first with probability $1/3$. John's personal "values" may reflect a disdain for fighting (soft John), and yet his rational adaptation to his circumstances lead him to behave in a way that is inconsistent with those "values."