# A Reformulation of a Criticism of The Intuitive Criterion and Forward Induction

George J. Mailath
University of Pennsylvania
gmailath@econ.upenn.edu

April 1988 (references updated July 2002)

The intuitive criterion of Kreps (see Cho and Kreps (1987, p. 202)) has been criticized by Stiglitz (see Cho and Kreps (1987, p. 203)), Mailath, Okuno-Fujiwara, and Postlewaite (1993) and van Damme (1989) for seeming inconsistencies in the way the reasoning is applied. Using the beer-quiche game as an example, this note recasts their criticism in a normal form argument which disputes the persuasiveness of the (naive) argument for not only the intuitive criterion, but also the requirement of robustness to elimination of never a weak best response (NWBR) strategies of Kohlberg and Mertens (1986) (a more general requirement which implies the intuitive criterion).

The argument for requiring robustness of "reasonable" equilibria to the elimination of NWBR strategies has a superficial similarity to the justification of elimination of weakly dominated strategies. Kohlberg and Mertens (1986, p. 1029) state that "[I]t seems natural to expect, based on first principles, that a strategically stable equilibrium must remain so after the deletion of a strategy which is an inferior response (at that equilibrium)." This criterion is commonly referred to as forward induction. While they do not expand on these first principles, the idea appears to be that, since the equilibrium is supposed to be common knowledge, a player will certainly not play a NWBR strategy. The other players know this and so do not need to allow for that possibility. Thus (the claim is) eliminating NWBR strategies will not destroy any reasonable equilibriums.[1]

Slightly abusing the terminology of Bernheim (1984) and Pearce (1984), I will say that a strategy profile can be rationalized from another strategy profile if there is a sequence of undominated best responses beginning at the second profile and ending at the first. Suppose a strategy profile has been isolated through iterated elimination of dominated and NWBR strategies (while the ideas are more general, it suffices in the

[1] The argument supporting elimination of weakly dominated strategies is almost identical, once admissibility is required—i.e., agents do not play weakly dominated strategies.

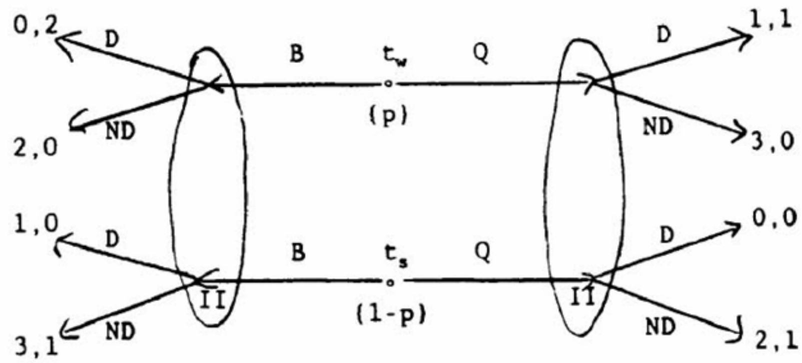


Figure 1: The Beer-Quiche game.

example to restrict attention to this case, and this is enough to make the point). I will argue that only if no element of the original component can be rationalized (in the original game) from this isolated strategy profile should the component be regarded as "failing forward induction." One reason for preferring this weaker version is a simple observation. While Nash equilibria of games obtained by deletion of dominated strategies are Nash equilibria of the original game, Nash equilibria of games obtained by deletion of NWBR strategies need not be equilibria of the original game. Thus it is necessary at some point to return to the original game.

The example is a version of the beer-quiche game of Kreps (see Cho and Kreps (1987, p. 183) for a story). The extensive form of the game is given in Figure 1 (this is the game from Cho and Kreps, except *II*'s payoff when $t_w$ chooses $B$ and *II* chooses $D$ is 2, rather than 1). As usual, *I*'s payoff is first and *II*'s payoff is second, and *I* moves first. Suppose $p < 0.5$.

Each player's strategy is a doubleton: the first (second) element of *I*'s strategy specifies an action, $Q$ or $B$, when $I$ is $t_w$ ($t_s$); the first (second) element of *II*'s strategy specifies an action, $D$ or $ND$, when $Q$ ($B$) is observed. The component of Nash equilibria yielding both types of $I$ choosing $Q$ has $I$ playing $(Q,Q)$ and *II* randomizing between $(ND, D)$ and $(ND, ND)$ with probability of at least 0.5 on $(ND, D)$. The intuitive criterion eliminates this outcome: If $I$ was of type $t_w$ then deviating to $B$ yields a payoff strictly less than 3, the equilibrium payoff, no matter how the deviation is interpreted. On the other hand, if $B$ is interpreted as coming from $t_s$, then $t_s$ prefers to deviate to $B$. Thus, *II* should interpret $B$ as a signal that $I$ is type $t_s$, and so the $(Q,Q)$ component fails the intuitive criterion.

But, the counterargument goes, since $t_s$ is better off by choosing $B$, $II$ should regard $Q$ as a sure signal that $I$ is in fact $t_w$ and so choose $D$ when $Q$. But then $t_w$ prefers $B$ to $Q$, which destroys the original contention that $B$ can be regarded as a signal that $I$ is $t_s$. This counterargument can be interpreted as a claim that the reasoning of the previous paragraph stopped at too early a stage. It is not an argument against all forms of forward induction, and in particular the version present above is not always vulnerable, as will be clear in the normal form presentation of the arguments.

Consider now the normal form elimination of strategies. Two values of $p$ will be considered, $p = 3/8$ and $p = 1/10$. The normal form for this game when $p = 3/8$ is (payoffs have been multiplied by 8):

| | | *II* | | | |
|---|---|---|---|---|---|
| | | $D, D$ | $D, ND$ | $ND, D$ | $ND, ND$ |
| | $Q, Q$ | $3, 3$ | $3, 3$ | $19, 5$ | $19, 5$ |
| $I$ | $Q, B$ | $8, 3$ | $18, 8$ | $14, 0$ | $24, 5$ |
| | $B, Q$ | $0, 6$ | $6, 0$ | $10, 11$ | $16, 5$ |
| | $B, B$ | $5, 6$ | $21, 5$ | $5, 6$ | $21, 5$ |

Observe first that $(B, Q)$ and $(B, B)$ are both NWBR in the $(Q, Q)$ component (both strategies yield less than or equal to 13/8 against any strategy of $II$'s in the component and the equilibrium outcome yields 19/8). The pruned game is

| | | *II* | | | |
|---|---|---|---|---|---|
| | | $D, D$ | $D, ND$ | $ND, D$ | $ND, ND$ |
| | $Q, Q$ | $3, 3$ | $3, 3$ | $19, 5$ | $19, 5$ |
| $I$ | $Q, B$ | $8, 3$ | $18, 8$ | $14, 0$ | $24, 5$ |

and in this game, the last strategy for $II$ dominates (weakly) the third strategy and so the $(Q, Q)$ is eliminated (the third strategy needed a nonzero weight but it has been eliminated). The intuitive criterion stops at this point, as does checking for the stability of the $(Q, Q)$ component. Consider, though, the pruned game that results from the deletion of $II$'s third strategy:

| | | *II* | | |
|---|---|---|---|---|
| | | $D, D$ | $D, ND$ | $ND, ND$ |
| | $Q, Q$ | $3, 3$ | $3, 3$ | $19, 5$ |
| $I$ | $Q, B$ | $8, 3$ | $18, 8$ | $24, 5$ |

In this game, $(Q, B)$ dominates $(Q, Q)$ and so $(D, ND)$ is the only reasonable strategy for $II$.

Now, $(D, ND)$ is $II$'s unique best response to $(Q, B)$ in the original game. However $I$'s unique best response to $(D, ND)$ in the original game is $(B, B)$ and $(ND, D)$ is a best response of $II$ to this strategy in the original game and so a possible strategy for $II$. Thus any profile in the $(Q, Q)$ component is rationalized from $((Q, B), (D, ND))$ and so the $(Q, Q)$ component does not fail forward induction.

This line of reasoning does not always save outcomes that have been eliminated by the intuitive criterion. Consider $p = 1/10$. The normal form in this case is:(payoffs have been multiplied by 10):

| | | $II$ | | | |
|---|---|---|---|---|---|
| | | $D, D$ | $D, ND$ | $ND, D$ | $ND, ND$ |
| | $Q, Q$ | $1, 1$ | $1, 1$ | $21, 9$ | $21, 9$ |
| $I$ | $Q, B$ | $10, 1$ | $28, 10$ | $12, 0$ | $30, 9$ |
| | $B, Q$ | $0, 2$ | $2, 0$ | $18, 11$ | $20, 9$ |
| | $B, B$ | $9, 2$ | $29, 9$ | $9, 2$ | $29, 9$ |

Again iterated deletion yields the profile $((Q, B), (D, ND))$. Returning to the original game, $(D, ND)$ is $II$'s unique best response to $(Q, B)$ while $(B, B)$ is $I$'s unique best response to $(D, ND)$. However $(ND, D)$ is *not* a best response for $II$ to $(B, B)$ and no profile in the $(Q, Q)$ component can be rationalized from $((Q, B), (D, ND))$. Thus the $(Q, Q)$ component fails forward induction and so is not reasonable.

The difference between the example for $p = 1/10$ and $3/8$ is that for $p = 1/10$, $(B, B)$ is an equilibrium outcome while for $p = 3/8$ it is not. In the latter case there is another equilibrium which has $I$ randomizing between $(B, B)$ and $(Q, B)$ with probability $5/6$ on the first strategy and $II$ randomizing between $(D, D)$ and $(D, ND)$ with probability $1/2$ on each strategy. The argument given here for the elimination of the $(Q, Q)$ component is certainly reminiscent of the undefeated concept of Mailath, Okuno-Fujiwara, and Postlewaite (1993), in that it seems to rely on the existence of a particular equilibrium. While the precise relationship is unknown (in particular their concept does not eliminate any equilibria in these examples as it is formally defined), something can be said which is suggestive. When their definition is relaxed as they mention on p. 13 then if p - 0.1, the (Q,Q) equilibrium is defeated by the (B,B) equilibrium. On the other hand, if p - 3/8, then the (Q,Q) equilibrium is undefeated.

## References


Bernheim, B. D. (1984): "Rationalizable Strategic Behavior," *Econometrica*, 52, 1007–1028.

Cho, I.-K., and D. Kreps (1987): "Signaling Games and Stable Equilibria," *Quarterly Journal of Economics*, 102, 179–221.

Kohlberg, E., and J.-F. Mertens (1986): "On the Strategic Stability of Equilibria," *Econometrica*, 54, 1003–1037.

Mailath, G. J., M. Okuno-Fujiwara, and A. Postlewaite (1993): "Belief based refinements in signaling games," *Journal of Economic Theory*, 60, 241–276.

Pearce, D. (1984): "Rationalizable strategic behavior and the problem of perfection," *Econometrica*, 52, 1029–50.

van Damme, E. (1989): "Stable Equilibria and Forward Induction," *Journal of Economic Theory*, 48, 476–496.