

HOW MUCH OF “FORWARD INDUCTION” IS IMPLIED BY “BACKWARD INDUCTION” AND “ORDINALITY”?

JOHN HILLAS

ABSTRACT. The examination of a number of examples suggests that there is an intimate connection between the concept of forward induction and that of backward induction. Following the literature on strategic stability, I examine (connected—even strongly connected) sets of (normal form perfect) equilibria. The kind of result that I am aiming at would claim that requiring some form of backward induction together with various forms of invariance of the solution to “inessential” aspects of the game would imply some form of forward induction. In fact, in all of the examples requiring that the set contain a subgame perfect equilibrium suffices as the backward induction requirement. It is however clear that such a requirement will not, in general, be sufficient and that one will need something such as the sets containing a quasi-perfect equilibrium. The solution is required to depend only on the reduced normal form of the game. However somewhat more invariance is required. A stronger form of this reduced normal form invariance is that the solution depend only on the best reply correspondence—or even the admissible best reply correspondence. An example is given that shows that such a strengthening of reduced normal form invariance is necessary for (at least one form of) the desired result. Even more is needed. One also requires that if two players never play in the same play of the game then the solution does not depend on whether they are considered as a single player or as two players. The form of the forward induction however does not appear to be at all delicate. In all of the examples the requirement of backward induction and invariance seems to imply something very close to the full strength of strategic stability, and this even in games in which it has been argued the such strength is “unintuitive.”

CONTENTS

1. Introduction	1
2. An Example of Kohlberg and Mertens	2
3. An Example of Cho and Kreps	3
4. Another Example	5
5. Mertens’ Example	6
6. More Invariance	8
7. Theorems?	8
8. Conclusion	9
References	10

1. INTRODUCTION

The concept of strategic stability introduced by Kohlberg and Mertens (1986) and redefined and developed by Mertens (1987,1989,1991a) aims to identify the self-enforcing outcomes of a game. Notice that this does not mean the same thing as identifying what will happen when the game is played, or even quite what could happen when the game is played. Rather it is about identifying the outcomes that are consistent with the interaction of the individual decision problems of the players. Or, in the negative, identifying the outcomes that can be ruled out because some player's individual incentives would lead him to deviate from that outcome.

The idea that the outcome of a game should be self-enforcing is quite old, being one of the most common justifications for Nash equilibrium. The conditions for Nash equilibrium are however only necessary conditions for an outcome to be self-enforcing. Selten (1965,1975) introduced the concepts of subgame perfect and (extensive form) perfect equilibrium in order to satisfy what he saw as additional requirements for an outcome to be self-enforcing. Kreps and Wilson (1982) and van Damme (1984) introduced the related concepts of sequential equilibrium and quasi-perfect equilibrium as answers to essentially the question Selten was addressing. The idea behind these concepts is that if some outcome is to be self-enforcing then the out of equilibrium behavior that supports that outcome should also be self-enforcing. This idea is generally known as backward induction.

Now, if one asks which outcomes are consistent with the interaction of the individual decision problems of the players then any change in the game that makes no essential change in those problems or in the manner in which they interact should not alter which outcomes are self-enforcing. This idea led Kohlberg and Mertens to require that the solution to a game depend only on the reduced normal form of the game. For the individual decision problems are essentially characterized by their reduced normal form and do not interact in any essential way beyond the interaction of the reduced normal forms.

Mertens (1987) extends and develops this idea under the term "ordinality." In particular he argues that the self-enforcing outcomes should depend only on the best reply correspondence, or even the admissible best reply correspondence. (The admissible best reply correspondence can be found from the best reply correspondence but not the reverse.) Mertens (1989,1991a) also adds other invariances, in particular arguing that, if two players never both move in the same play of the game, then it should not matter whether they are considered as one player or as two. In both cases the argument is the same as for reduced normal form invariance; the changes in the game do not change, in any essential way, either the individual decision problems or the way in which they interact.

Already "backward induction" and reduced normal form invariance are inconsistent for a single valued solution concept. This makes it necessary to look at sets of equilibria and to interpret "backward induction" to mean that the set *contains* a "backward induction" (e.g., sequential) equilibrium. This is why I have been referring imprecisely to self-enforcing outcomes, rather than self-enforcing strategy vectors. The precise status and meaning of such sets is, to my mind, not currently completely understood. At a minimum they should be connected, or even perhaps strongly connected as required in the definition of Mertens. This issue will not be developed in this paper.

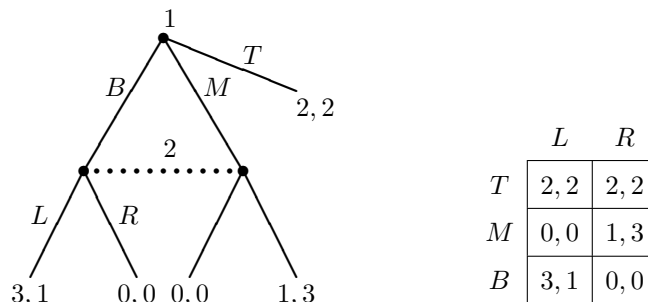


FIGURE 1

Another idea that is discussed by Kohlberg and Mertens and that is present in the informal motivation in their paper is the idea of “forward induction.” The precise status of this concept is not clear in their paper. They do not list “forward induction” as one of the requirements for a solution but it seems to be important in the motivation and stable sets do, in fact, satisfy their concept of forward induction. The property that Kohlberg and Mertens call forward induction is the following: “A stable set contains a stable set of any game obtained by the deletion of a strategy which is an inferior response in all the equilibria of the set.” (Kohlberg and Mertens, p. 1029)

This is obviously a strong property. For example, it means that if there are two strategies that are both inferior responses then when the more preferred of these strategies is deleted and the less preferred kept the solution should remain stable. Even before this full strength there are those who have argued against this kind of requirement. Cho and Kreps (1987) show how a series of stronger implementations of the idea of forward induction refine the set of sequential equilibria in signaling games. They argue that the relatively strong implementations are quite unintuitive.

In the following sections I examine a number of examples of games to which the idea of forward induction has been applied in the literature. In these examples the combination of forward induction and reduced normal form invariance are sufficient to eliminate the sets of equilibria that are not stable. A number of the examples do depend on the fact that in cases in which we could treat agents as either different players or the same player we treat the agents as the same player.

2. AN EXAMPLE OF KOHLBERG AND MERTENS

In this section I examine a classic game—given here in Figure 1—from Kohlberg and Mertens (1986). This game was important in motivating the idea of “forward induction.” There are equilibria in which player 1 plays *T*. The argument that has been given against such equilibria is as follows: If player 2 is called on to move he does not see whether player 1 has played *M* or *B*. However he does know that if player 1 chooses *M* the most that player 1 can obtain is 1 while player 1 could have obtained 2 with certainty by choosing *T*. On the other hand if player 1 can convince player 2 that he has indeed played strategy *B*—so that player 2 will choose *L*—then player 1 gains by playing *B*. Thus when player 2 sees that player 1 has played either *M* or *B* player 2 should believe that player 1 has actually played *B*. This will lead player 2 to play *L* and so player 1 should not play *T* but rather *B*.

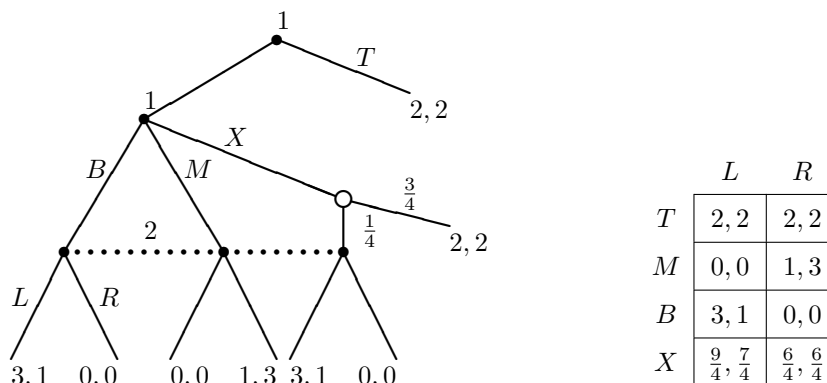


FIGURE 2

This reasoning and variants of it are what is meant by “forward induction”. One is either convinced by such reasoning or not. I myself find it fairly compelling. In this paper however I wish to examine how much of the results that follow from this kind of reasoning can be obtained from two other kinds of properties: invariance to “inessential” changes in the game; and backward induction. I shall become more precise about these properties as the paper progresses. For this example all that is needed is invariance to the reduced normal form and that the solution contain a subgame perfect equilibrium.

The game shown in Figure 2 has the same reduced normal form as the original game of Figure 1. Nevertheless one does not require any kind of forward induction argument to argue against the equilibria in which player 1 plays T in this game. Subgame perfection suffices. It is straightforward to see that (B, L) is the unique equilibrium of the subgame, and so (B, L) is the unique subgame perfect—and hence unique sequential—equilibrium of the game.

I think that most game theorists—at least most of those who are at all interested in refinements—agreed that the equilibria of this game in which player 1 plays T are “unreasonable”. (Substitute another term for unreasonable if you wish.) So it is not surprising that other justifications for getting rid of these equilibria can be found. I shall show in the following two sections that the same reasoning serves to eliminate equilibria in games in which Cho and Kreps had argued that the forward induction type arguments were unintuitively strong.

3. AN EXAMPLE OF CHO AND KREPS

In this section and the next I shall examine two games from Cho and Kreps (1987). For both games Cho and Kreps gave arguments of the kind I am calling forward induction that eliminated some of the equilibria of these games. They expressed misgivings about the strength of these criteria.

The game considered in this section is given by Kreps to illustrate the use of the criterion they call “Never a Weak Best Response.” That criterion requires that when a strategy that is an inferior response at all of the sequential equilibria in the set under consideration is deleted the set should, in the game that results, continue to contain a sequential equilibrium. The game suggests to Cho and Kreps that this criterion is unintuitively strong. The game and the normal form of the game are

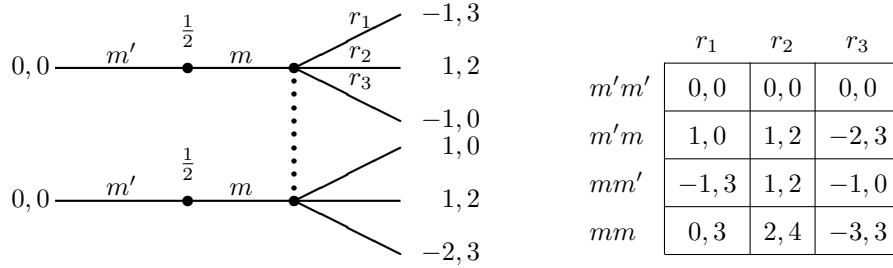


FIGURE 3

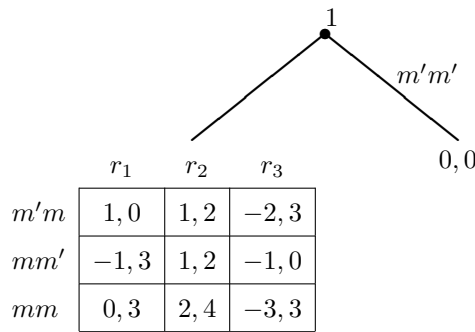


FIGURE 4

given here in Figure 3. (The payoffs in the normal form have all been multiplied by 2 to eliminate fractions.)

There are two disjoint sets of equilibria in this game $\{(0, 0, 0, 1), (0, 1, 0)\}$ and $\{(1, 0, 0, 0), (x, y, 1 - x - y) \mid x + y \leq \frac{2}{3} \text{ and } y \leq \frac{1}{2}\}$. In the second set the normal form perfect equilibria are $\{(1, 0, 0, 0), (0, y, 1 - y) \mid y \leq \frac{1}{2}\}$. We see already in the normal form that none of these equilibria are proper, and so the set is not stable (in the sense of Mertens' new definition or the definition of Hillas (1990)). In fact this set is not stable even in the sense of the original definition. The perturbation $(\delta^2, \delta^2, \delta, \delta^2), (\text{anything})$ gives a game with no equilibria close to the second set.

As in the previous section the same result follows simply from reduced normal form invariance and backward induction. In fact, the full strength of these requirements are not needed in this game. Only invariance to the elimination of duplicate pure strategies in the normal form and subgame perfection are needed. Figure 4 gives a game with the same normal form as the game of Figure 3. The subgame has a unique equilibrium $((0, 0, 1), (0, 1, 0))$ and so the game has a unique subgame perfect equilibrium $((0, 0, 0, 1), (0, 1, 0))$. Again the requirements of invariance and backward induction lead to the full strength of stability.

In order to argue as I have it is necessary to think of the two types of Player 1 as truly being the *same* player at two different information sets. The game to which the requirement of backward induction is applied certainly does not have the same reduced normal form as the original game if the different types of player 1 are treated as actually being *different* players. For the moment I shall simply state this as a possible caveat and return to the point later.

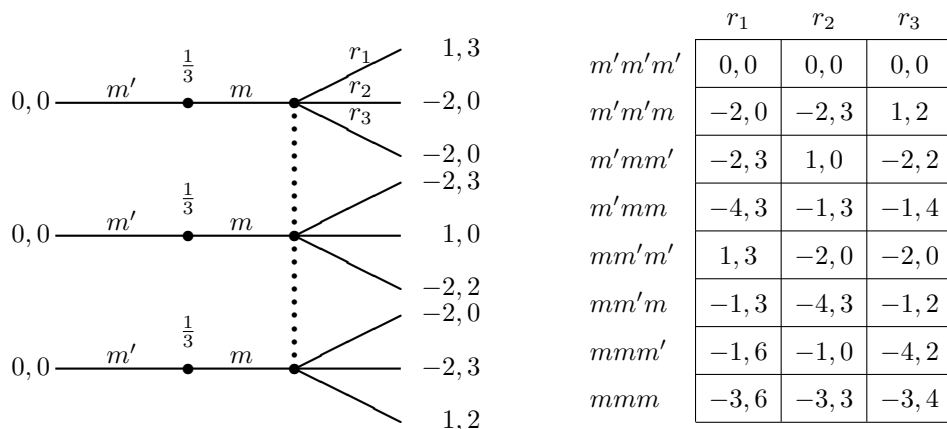


FIGURE 5

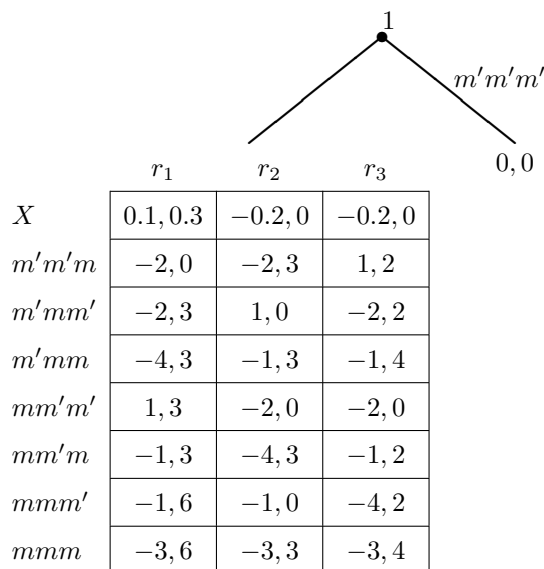


FIGURE 6

4. ANOTHER EXAMPLE

In this section I examine another example from Cho and Kreps. They argue that this example shows the unintuitiveness of the “full strength of stability”. In this game not even the strongest implementations of “forward induction” short of stability itself eliminate the equilibrium under consideration. However, as in the previous examples, reduced normal form invariance and backward induction do eliminate the set that is not strategically stable.

The game and its normal form are given here in Figure 5. (The payoffs in the normal form are all multiplied by three to eliminate fractions.) I shall focus, as do Cho and Kreps, on the set of equilibria in which all types of player 1 play m' ,

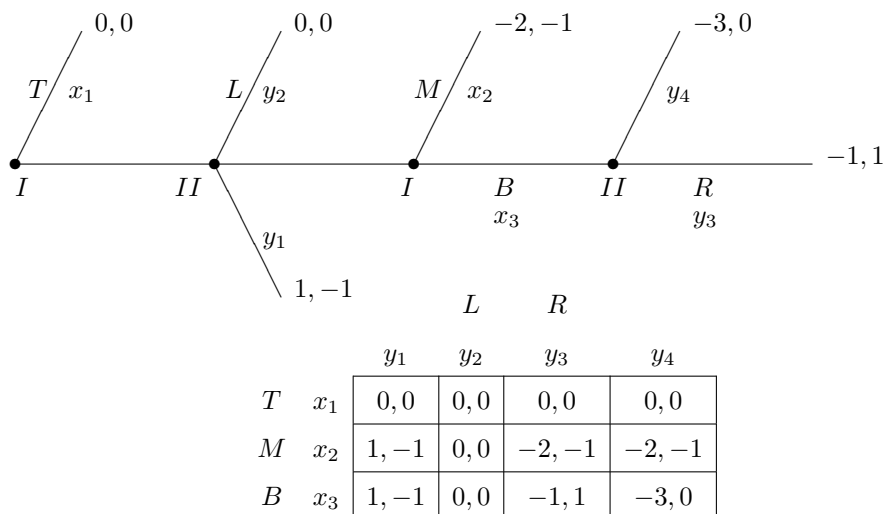


FIGURE 7

i.e., on the set $\{((1, 0, 0, 0, 0, 0, 0), (x, y, 1 - x - y)) \mid x + y \geq \frac{1}{3}, x \leq \frac{2}{3}, y \leq \frac{2}{3}\}$. These equilibria are all normal form perfect—and so in the signaling game are all sequential. The perturbation $((\delta^2, \delta^2, \delta^2, \delta^2, \delta, \delta^2, \delta^2, \delta^2), (\text{anything}))$ defines a game with no equilibria close to the set under consideration.

In this game the full strength of reduced normal form invariance is required to make the argument. In the example of the previous section I showed that in a game with the same normal form as the original game there were no subgame perfect equilibria in the set of equilibria being considered. In the current game there is always a sequential equilibrium in the set for any game with the same normal form. Recall the result of van Damme (1984) and of Kohlberg and Mertens (1986) that a proper equilibrium of a normal form game is sequential in any game with that normal form. The equilibrium $((1, 0, 0, 0, 0, 0, 0), (\frac{1}{3}, \frac{1}{3}, \frac{1}{3}))$ is proper.

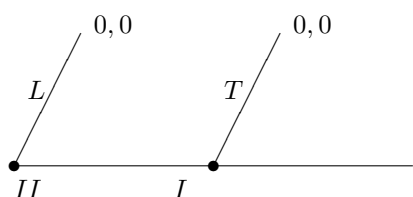
However, as in our first example, reduced normal form invariance is more. Suppose the mixed strategy $(0.9, 0, 0, 0, 0.1, 0, 0, 0)$ is added as a new pure strategy for player 1 and call it X . The game shown in Figure 6 has such a normal form. Thus, according to the requirement of reduced normal form invariance, this game should have the same solution as the original game. However in this game the subgame has a unique equilibrium $((0, 0, 0, 1, 0, 0, 0), (1, 0, 0))$ and so the game has a unique subgame perfect equilibrium $((0, 0, 0, 0, 1, 0, 0, 0), (1, 0, 0))$. Since this is not in the set under consideration that set does not satisfy our requirements.

5. MERTENS' EXAMPLE

In this section I examine a remarkable game recently given by Mertens (1991b). This is a game with perfect information and a unique subgame perfect equilibrium—though the game is not generic. Moreover there are completely normal form arguments that support the unique subgame perfect equilibrium. In spite this the unique strategically stable set of equilibria consists of all the admissible equilibria, a set containing much besides the subgame perfect equilibrium. The game is given here in Figure 7.

	y_1	L	R	y_4	Y
T	0, 0	0, 0	0, 0	0, 0	0, 0
M	1, -1	0, 0	-2, -1	-2, -1	0.02, -0.14
B	1, -1	0, 0	-1, 1	-3, 0	0.02, -0.1
X	0.1, -0.1	0, 0	-0.2, -0.1	-0.2, -0.1	0.002, -0.014

FIGURE 8A



	y_1	R	y_4	Y
M	1, -1	-2, -1	-2, -1	0.02, -0.14
B	1, -1	-1, 1	-3, 0	0.02, -0.1
X	0.1, -0.1	-0.2, -0.1	-0.2, -0.1	0.002, -0.014

FIGURE 8B

The unique subgame perfect equilibrium of this game is (T, R) . Thus this is also the unique proper equilibrium—though the uniqueness of the proper equilibrium does involve an identification of duplicate strategies that is not required for the uniqueness of the subgame perfect equilibrium—and so (T, R) is a sequential equilibrium of any extensive form game having this normal form. Nevertheless there are games having the same *reduced* normal form as this game for which (T, R) is not sequential, and moreover the outcome of the game is not the same as the outcome from (T, R) .

Suppose that the mixtures $X = 0.9T + 0.1M$ and $Y = 0.86L + 0.1y_1 + 0.04y_4$ are added as new pure strategies. This results in the normal form game given in Figure 8a. An extensive form game having this normal form is given in Figure 8b. It is straightforward to see that the unique equilibrium of the simultaneous move subgame is (M, Y) in which the payoff to Player I is strictly positive (0.02) and so in any subgame perfect equilibrium Player I chooses to play “in”—i.e., M , or B , or X —at his first move. Moreover in the equilibrium of the subgame the payoff to Player 2 is strictly negative (-0.14) and so in any subgame perfect equilibrium Player II chooses to play L at his first move. Thus the unique subgame perfect equilibrium of this game is (M, L) .

This equilibrium is far from the unique backward induction equilibrium of the original game. Moreover its outcome is also different. Thus, if one accepts the arguments that the solution(s) of a game should depend only on the reduced normal form of the game and that any solution to a game should contain a “backward induction” equilibrium, then one is forced to the conclusion that in this game the solution must contain much besides the subgame perfect equilibrium of the original presentation of the game.

This gives an additional independent support to Mertens’ construction in the original extensive form of a manner in which the players might reason that would

support equilibria other than the subgame perfect equilibrium of that extensive form game.

6. MORE INVARIANCE

As I commented earlier the argument in some of the examples—in particular in the two games of Cho and Kreps—depended on treating the two types of one of the players as being the same player. If the games had initially been presented as three player games the arguments of the earlier sections would apparently not apply. Also, it would not change the essential features of the original extensive form game in Section 5 if Player II's payoff following T is changed to 10, say. However, in this case it would not be possible to change the order of Players I and II's first two moves as is done in the extensive form game given in Figure 8b.

The answer to both of these apparent difficulties is to increase the invariances that a solution be required to satisfy. In any case, though I shall not make the arguments here, it is reasonable that an answer to the question, “What are the self-enforcing solutions in this game?” should exhibit such invariances. I address the second concern first, since the invariance in this case is part of what Mertens means by “ordinality”.

Mertens (1987, 1989) argues that the solution to a game should not only depend only on the reduced normal form of the game in question, but also depend only on the (mixed) best reply structure of the game. I shall not attempt to argue for this view here, though I certainly agree with Mertens. If one accepts this view then one merely needs to note that the change of Player II's payoff to 10 in the example under consideration does not alter the best reply structure and so, in making the argument, one would simply start by appealing to the invariance to changes that leave the best reply structure unchanged and then simply change the 10 back to 0.

The answer to the first difficulty is similar. One could argue, as Mertens does (and, again, I agree with him), that if, in an extensive form game, two players never both move in the same play of a game then it should not matter for the solution whether those players are treated as one player or as two. Again I do not intend to argue for this assumption here. If one accepts this assumption then it certainly solves the difficulty. One is free to think of the players in question in the example as a single player.

7. THEOREMS?

In the examples I have considered above it was not necessary to appeal to any form of “backward induction” stronger than subgame perfection. It will not be true in general that the various invariances proposed above and subgame perfection will imply any forward induction property. For a given normal form game it would be possible to add additional “dummy players” who have a choice to make, but whose choice does not affect the players who are “really” in the game and whose outcome for the player making the choice depends on the actions of the “real” players. With enough players appropriately added neither subgame perfection nor sequentiality will have any implications whatsoever. One might perhaps try to give a definition of precisely what it means to be a “dummy” but this seems rather difficult.

Rather I propose to strengthen the requirement of “backward induction”. The examples seem to indicate that this strengthening will be needed only in the case of pathologies such as the one described in the previous paragraph, but I don't see

how this could be made precise. The appropriate strengthening seems to me to be the concept of quasi-perfect equilibria of van Damme (1984). (See the discussion in Mertens (1991b) for the reason why one would not wish to use the concept of (extensive-form) perfect equilibria of Selten (1975). Also see Reny (1988) for an argument that all of the backward induction properties I have used in this paper are too strong.)

As I commented in Section 2, any solution satisfying both some form of backward induction at least as strong as subgame perfection and reduced normal form invariance must necessarily be multivalued. In dealing with such solutions one wishes to remain reasonably close to a single valued solution. Kohlberg and Mertens did so by imposing a minimality condition while Mertens in his redefinition simply restricts himself to sets of equilibria that are connected on the interior of the space of perturbations. Such sets he calls strongly connected. He also requires that the sets not suddenly become bigger on the boundary of the space of perturbations. Among other implications of this is that the sets will consist of only normal form perfect equilibria. I shall call a set satisfying these conditions a strongly connected set of normal form perfect equilibria.

Thus I intend to examine the concept of invariant quasi-perfect sets of equilibria.

DEFINITION: An *invariant quasi-perfect set of equilibria* of a game is a strongly connected set of normal form perfect equilibria of that game such that any extensive form game that is equivalent to this game has the a quasi-perfect equilibrium in the set, where equivalent means that the normal form of one can be obtained from the normal form of the other by

- (i) changes to the payoffs that leave the best reply structure unchanged,
- (ii) the adding or deletion of mixtures of other strategies, or
- (iii) changing the treatment of two players that never both move in the same play of the game.

Since a stable set of equilibria as defined by Mertens satisfies these conditions invariant quasi-perfect sets always exist. I shall investigate whether such a solution satisfies the property that Kohlberg and Mertens call forward induction whose definition was quoted in the Introduction. It would also be interesting to know how such sets relate to the various definitions of stable sets defined in Kohlberg and Mertens (1986), Mertens (1989, 1991a), and Hillas (1990). It is already clear that stable sets in the sense of Kohlberg and Mertens may not be invariant quasi-perfect sets (see the example of Gul in the paper of Kohlberg and Mertens) and that stable sets in the sense of Mertens' redefinition are necessarily invariant quasi-perfect sets.

8. CONCLUSION

There is by now a large literature on the refinement of Nash equilibrium. The paper of Kohlberg and Mertens on strategic stability in some ways "raised the stakes" in this field. While neither it nor the papers of Mertens(1987, 1989, 1991a) that followed were formally axiomatic, in the sense that much of cooperative game theory is, they were, at least in my reading, very much in that spirit. Kohlberg and Mertens motivate their definition with a list of requirements that they argue a solution concept should satisfy. While none of the solutions defined in that paper actually satisfy these requirements and no current definition of stability has been

shown to be characterized by these requirements this was already (again, in my opinion) a major break from the previous literature.

The papers of Mertens extended the list of required properties, found a definition that did satisfy the requirements (though in certain minor ways only in a qualified manner), and tied the details of the definition much more intimately to the requirements. (Of course, even this treatment is not formally axiomatic.)

This paper is complementary to that work. It looks somewhat more directly at the requirements themselves. In particular it looks at the relation between “backward induction” and various “invariance” requirements on the one hand and “forward induction” on the other. A number of examples suggested that forward induction is an implication of the other two requirements. Section 6 provided some ideas on how such a result might be formalized.

REFERENCES

- Cho, In-Koo, and David M. Kreps (1987): “Signaling Games and Stable Equilibria,” *Quarterly Journal of Economics*, 102, 179–221.
- Hillas, John (1990): “On the Definition of the Strategic Stability of Equilibria,” *Econometrica*, 58, 1365–1390.
- Kohlberg, Elon and Jean-François Mertens (1986): “On the Strategic Stability of Equilibria,” *Econometrica*, 54, 1003–1038.
- Mertens, Jean-François (1987): “Ordinality in Non Cooperative Games,” CORE Discussion Paper 8728, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
- Mertens, Jean-François (1989): “Stable Equilibria—A Reformulation, Part I: Definition and Basic Properties,” *Mathematics of Operations Research*, 14, 575–624.
- Mertens, Jean-François (1991a): “Stable Equilibria—A Reformulation, Part II: Discussion of the Definition, and Further Results,” *Mathematics of Operations Research*, 16, 694–753.
- Mertens, Jean-François (1991b): “Two Examples on Strategic Equilibria,” Discussion Paper DP-91-27, Institute for Decision Sciences, SUNY at Stony Brook.
- Nash, John (1950): “Equilibrium Points in N -Person Games,” *Proceedings of the National Academy of Science*, 36, 48–49.
- Nash, John (1951): “Non-Cooperative Games,” *Annals of Mathematics*, 54, 286–295.
- Reny, Philip J. (1992): “Backward Induction, Normal Form Perfection and Explorable Equilibria,” *Econometrica*, 60, 627–649.
- Selten, Reinhard (1965): “Spieltheoretische Behandlung eines Oligopolmodells mit Nachfrageträgheit,” *Zeitschrift für die gesamte Staatswissenschaft*, 121, 301–324.
- Selten, Reinhard (1975): “Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games,” *International Journal of Game Theory*, 4, 25–55.
- Van Damme, Eric (1984): “A Relation Between Perfect Equilibria in Extensive Form Games and Proper Equilibria in Normal Form Games,” *International Journal of Game Theory*, 13, 1–13.

DEPARTMENT OF ECONOMICS, UNIVERSITY OF AUCKLAND
 E-mail address: j.hillas@auckland.ac.nz