INSTITUTE OF MATHEMATICAL ECONOMICS

WORKING PAPERS

No. 322

Can and Should the Nash Program be Looked at as a Part of Mechanism Theory?

by

Walter Trockel

April 2001

University of Bielefeld

33501 Bielefeld, Germany

Can and should the Nash program be looked at as a part of mechanism theory?

Walter Trockel
IMW, Bielefeld University
email:WTROCKEL@WIWI.UNI-BIELEFELD.DE

April 2001

Abstract

The present article presents and discusses the recent controversy about the possibility and meaning of relating the game theoretic Nash program to mechanism theory. The non-cooperative foundation of the Nash bargaining solution is used as an example to shed light on the formal relation between traditional non-cooperative support of cooperative solutions and mechanism theoretic implementation of social choice rules. The technical possibility of regarding the Nash program as a part of mechanism theory is taken as a starting point for discussing the meaning of implementation, "good" mechanisms and extensions in the spirit of Hurwicz's (1994) "genuine" implementation.

Content

1 The Nash Program

The Nash Program is a research agenda whose goal it is to provide a non-cooperative equilibrium foundation for axiomatically defined solutions of cooperative games. This program was initiated by John Nash in his seminal papers *Non-cooperative Games* in the *Annals of Mathematics*, 1951, and *Two-Person Cooperative Games* in *Econometrica*, 1953. The term *Nash Program* introduced by Binmore (1987)(see also Binmore, 1997) had been used according to Reinhard Selten (1999) much earlier in a lecture by Robert Aumann. The original passages due to Nash that built the basis for this terming are in fact quite short.

In Nash (1951) one reads:

"A less obvious type of application (of non-cooperative games) is to the study of cooperative games. By a cooperative game we mean a situation involving a set of players, pure strategies, and payoffs as usual; but with the assumption that the players can and will collaborate as they do in the von Neumann and Morgenstern theory. This means the players may communicate and form coalitions which will be enforced by an umpire. It is unnecessarily restrictive, however, to assume any transferability or even comparability of the pay-offs [which should be in utility units] to different players. Any desired transferability can be put into the game itself instead of assuming it possible in the extra-game collaboration.

The writer has developed a "dynamical" approach to the study of cooperative games based upon reduction to non-cooperative form. One proceeds by constructing a model of the pre-play negotiation so that the steps of negotiation become moves in a larger non-cooperative game [which will have an infinity of pure strategies] describing the total situation. This larger game is then treated in terms of the theory of this paper [extended to infinite games] and if values are obtained they are taken as the values of the cooperative game. Thus the problem of analyzing a cooperative game becomes the problem of obtaining a suitable, and convincing, non-cooperative model for the negotiation.

The writer has, by such a treatment, obtained values for all finite two-person cooperative games, and some special n-person games".

The work Nash is referring to in his last sentence of the above passage is Nash (1953) where he writes:

"We give two independent derivations of our solution of the two-person cooperative game. In the first, the cooperative game is reduced to a non-cooperative game. To do this, one makes the players' steps of negotiation in the cooperative game become moves in the non-cooperative model. Of course, one cannot represent all possible bargaining devices as moves in the non-cooperative game. The negotiation process must be formalized and restricted, but in such a way that each participant is still able to utilize all the essential strengths of

his position.

The second approach is by the axiomatic method. One states as axioms several properties that it would seem natural for the solution to have and then one discovers that the axioms actually determine the solution uniquely. The two approaches to the problem, via the negotiation model or via the axioms, are complementary; each helps to justify and clarify the other."

The Nash program tries to link two different ways of solving games. The first one is non-cooperative. No agreements on outcomes are enforceable. Hence players are totally dependent on their own strategic actions. They try to find out what is best given, the other players are rational and do the same. In this context the Nash equilibrium describes a stable strategy profile where nobody would have an interest to unilaterally deviate. Nevertheless there is an implicit institutional context. The strategy sets define implicitly what choices are not allowed, those outside the strategy sets. The payoff functions reflect which strategies in the interplay with others' strategies are better or worse. It is not said explicitly who grants payoffs and how the physical process of paying them out is organized. But there is some juridical context with some enforcement power taken for granted. There is no interpersonal comparison of payoffs involved in the determination of good strategies. Each player only compares his different strategies contingent on the other players' different strategy choices. As applications in oligopoly show, institutional restrictions of social or economic scenarios are mapped into strategy sets and payoff functions, thereby lending them an institutional interpretation. Yet, totally different scenarios may considerably be modelled by the same non-cooperative game, say in strategic form. This demonstrates clearly the purely payoff based evaluation of games. Payoffs usually are interpreted as reflecting monetary or utility payments. Associated physical states or allocations orrur only in applications and may be different in distinct applications of the same game.

The second way to solve a game is the cooperative one via axioms as first advocated by Nash (1953). Again the legal framework is only implicit. Yet, now not only obidience to the rules is assumed to be enforceable but even contracts. Mutual gains are in reach now as it becomes possible by signing a contract to commit himself to certain behavior. In this context it is the specific payoff configuration which is of interest rather than the strategy profile that would generate it. In this framework it is reasonable, therefore, to neglect the strategic options and concentrate on the feasible payoff configurations or utility allocations on which the players possibly could agree by signing a contract. Again the formal model does not specify the process by which physical execution of a contract is performed. Again it is the payoff space rather than some underlying social scenario on which the interest rests except in applications of game theory.

In contrast to the non-cooperative approach now players **are** interested in what other players receive. Although utilities or payoff units for different players are in general not considered comparable typically there are tradeoffs that count. The axioms that are fundamental in this context reflect ideas of fairness, equity, justness that do not play a role in the non-cooperative model. But a process of negotiation with the goal to find an agreement makes it necessary for each player to somehow judge the coplayers' payoffs. But the axioms are in a purely welfaristic context. If very different underlying models lead to the same cooperative game in coalitional form it is only the solution in terms of payoff vectors that is relevant. And this determines in any application what underlying social or physical state is distinguished. It becomes irrelevant in the axiomatic cooperative approach which are the institutional details. Important are only the feasible utility allocations.

Now, why could it be interesting to have a non-cooperative strategic game and a cooperative game in coalitional form distinguishing via its equilibrium or solution, respectively, the same payoff vector? According to Nash the answer is that each approach *"helps to justify and clarify the other"*.

The equality of payoffs in both approaches seems to indicate that the institutional specifities represented by the strategic model are not so restrictive as to prevent the cooperative solution. Also the payoff function appears then to reflect in an adequate way the different axioms. On the other hand payoff combinations not adequate under the solution concept cannot be strategically stable. So the equivalence of both approaches seems to indicate that the strategic model from the point of view of social desirability is restrictive enough but not too restrictive. This abstract relation has different consequences if one is in one of the two different enforceability contexts. If we cannot enforce contracts the equivalence of two approaches means that this is not a real drawback, as we can reach the same via rational strategic interaction (at least in situations of games with a unique equilibrium). If, on the other hand, we are in a world where contracts are enforceable, we may use the equivalence of a suitable strategic approach as additional arguments for the payoff vectors distinguished by the solution. Therefore, results in the Nash program give players valuable insights into the interrelation between institutionally determined non-cooperative strategic interaction and social desirability based on welfaristic evaluation.

There is not, however, any focus on *decentralization* in the context of the Nash program simply because there is no entity like a center or planner. There are just players.

Nash's own first contribution to the Nash Program (1953) consists in his analysis of a game, the *demand game* and the so called *smoothed demand game* where he looked at the limiting behavior of non-cooperative equilibria of a sequence of smoothed versions of the demand game. Here the amount of smoothing approaches zero, and, hence the sequence approximates the demand game. While the original "simple" demand game has a continuum of equilibria, a fact which makes it useless for a non-cooperative foundation of the Nash solution, Nash argued that the Nash solution was the only necessary limit of equilibria of the smoothed games. Rigorous analyses for his procedure have been provided much later by Binmore (1987), van Damme (1987) and Osborne and Rubinstein (1990).

These passages make it quite clear that Nash changed slightly his stress between 1951 and 1953. While the first quotation lends more support to the interpretation of his main goal as a non-cooperative solution of cooperative games before and without having any cooperative solution for which non-cooperative foundation has to be provided, the second later passage argues that both approaches are equally valuable. In fact here it is the axiomatic cooperative solution which confirms the earlier non-cooperatively derived solution. I tend to interpret Nash's point of view as a dual one where the cooperative and the non-cooperative approaches mutually support each other.

Unfortunately, the non-cooperative approach of Nash to the bargaining problem failed to be fully successful. Nash did not provide a "suitable and convincing non-cooperative model" that supports the Nash solution. In the simple demand game the multiplicity of equilibria causes the failure of support, in the smoothed game approach there is not *one* game the equilibrium of which provides the support rather a sequence of games where a limit of equilibria provides the support. But the sequence cannot be played. What is provided is a distinguished role of the Nash solution among the infinity of equilibria of the simple demand game. In a preplay communication previous to playing the non-cooperative simple demand game players might be able to agree on the *focal point* role of this particular equilibrium. Hence Nash's analysis might be seen as a coordination device. Yet, one might argue that the hint to the symmetry property of the symmetric Nash solution could be at least as effective. Binmore and Dasgupta (1987) argue that it is a stability property of the symmetric Nash solution that distinguishes it among all the equilibria of the simple demand game. They even compare the slightly disturbed smooth versions with small trembles in agents' information and attribute to Nash that "he anticipated the essence of Selten's (1975) notion of a "'trembling-hand' equilibrium". However, there is hardly a sound basis for this interpretation. The focal or salient stability property of the Nash solution as a Nash equilibrium of the simple demand game is in van Damme's (1987) terminology its existence as a unique "H-essential" equilibrium of the demand game. The "H" refers to the class of perturbations of the demand game that are considered. A closer look at this class H in van Damme's analysis or at the analogous treatments of Binmore (1987) or of Osborne and Rubinstein (1990) reveals that the symmetry of the (symmetric) Nash solution is put in already into the class H. If we rename van Damme's H by $H_{1/2,1/2}$ it is easy to see that for any $\alpha \in (0,1)$ one can construct a class $H_{\alpha,1-\alpha}$ in a perfectly analogous way. The unique $H_{\alpha,1-\alpha}$-essential equilibrium of the simple demand game turns out to be the asymmetric Nash solution that maximizes $x_1^{\alpha}x_2^{1-\alpha}$ on the set of feasible payoff allocations.

In this sense either **each** of the equilibria of Nash's simple demand game is essential (for some α) or **none** (for all α). It is not any kind of stability that distinguishes the symmetric one from all the asymmetric Nash solutions - it is just its symmetry. This observation shows very clearly the difference of this approach from Selten's trembling hand perfectness. The latter one is a property of certain equilibria that distinguishes them inherently due to the structure of the considered game rather than due to distortions taylored to single out some pre-specified equilibrium.

A second quite different approximate non-cooperative support for the Nash solution is provided by Rubinstein's (1982) model of sequential alternate offers bargaining. Binmore, Rubinstein and Wolinsky (1986) showed in two different models with discounted time that the weaker the discounting is the more closely approximates the subgame perfect Nash equilibrium an asymmetric Nash bargaining solution. Only if subjective probabilities of breakdown of negotiations or the lengths of reaction times to the opponents' proposals are symmetric it is the symmetric Nash solution which is approximately supported. Again, in the frictionless limit model one does not get support of the Nash solution by a unique equilibrium. Rather every individually rational payoff vector corresponds to some subgame perfect equilibrium.

An exact support rather than only an approximate one of the Nash solution is due to Howard (1992). He proposes a fairly complex 10 stages extensive form game whose unique subgame perfect equilibrium payoff vector coincides with the bargaining solution.

There are several contributions to the Nash program for other solutions. Of particular interest in our context are support results for the Kalai-Smorodinsky solution (cf. Crawford (1978), Haake (2000), Moulin (1984), Trockel (1999b)), which is the most popular alternative for the Nash bargaining solution.

Like in Rubinstein's model and in contrast to Nash framework Howard's game is based on underlying outcome space. Here this is set of lotteries over some finite set on which players have utility functions. Although the analysis of the game can be performed without explicit consideration of the outcome space it is this underlying structure that allows it to look at the outcome associated with a subgame perfect equilibrium and thereby interpret Howard's support result as a mechanism theoretic implementation of some Nash social choice rule in subgame perfect equilibrium. The fact that such an underlying outcome space is not easily at hand in the purely welfaristic framework of Nash's axiomatic bargaining model causes the problem of extending a support result in this framework to a proper implementation result.

2 Mechanism Theory

The Nash Program is concerned with providing a justification for a certain payoff vector simultaneously by a solution concept for cooperative games and by an equilibrium of some non-cooperative game. The only actors playing a role are the players of both types of games.

In Mechanism Theory the situation is fundamentally different. A planner or designer thinks about the problem how to induce an arbitrary population of agents in a society to jointly realize a *social state* that, given the agents' preferences, is considered socially desirable. Lacking information about agents' preferences and sufficient enforcement power he aims to design the rules for strategic interaction, i.e. a *game form*, in such a way that any possible population of agents by adopting equilibrium behavior of the game that is induced by the game form together with agents' utility functions on social states realizes a state that is desirable for that specific society.

I want to follow at this point Matthew Jackson (1999) and stress a distinction between two parts of mechanism theory.

One is *mechanism design* where the interest in incentive compatibility leads to the question whether a certain (desirable) outcome can be induced as an equilibrium of some game form. It is the stability property inherent in an equilibrium that is looked for. This property is not lost if other equilibria exist, even if they do not induce desirable outcomes. In that context it appears acceptable to restrict to direct games and truthful implementation. The other part is *implementation theory.* Here indirect game forms are used but all equilibria are of concern. A mechanism is considered acceptable in this framework only if all equilibria induce desirable outcomes.

It is this second aspect of mechanism theory to which I shall relate the Nash program by showing how support results that provide a non-cooperative foundation may be extended to proper implementation results.

For this purpose we need formal definitions of a *social choice rule*, of a *game form*, and of its *implementation* in Nash equilibrium.

Let $I = \{1, ..., n\}$ be the set of players' positions, A be some non-empty set, called *outcome space* and M_i sets of possible messages m_i among which a player in position $i \in I$ may choose.

The outcome space represents all possible states for a n-person society. In applications it may be a set of allocions in an economy, a set of candidates in a voting context or a set of lotteries over monetary prizes. The only formal requirement for an outcome space A is that it is some non-empty set.

Let $U_i, i = 1, ..., n$ be non-empty sets of utility functions $u_i : A \longrightarrow R$. Let $U \subset U_1 \times ... \times U_n$ be the set of admissible profiles of utility functions. In the case of $U = U_1 \times ... \times U_n$ we speak of an *unrestricted domain* of utility function profiles. A correpondence $F : U \Longrightarrow A$ is called a *social choice rule.* If F is singleton-valued with $F(u) = \{f(u)\}$ we call f, or by slight abuse of notation also F, a *social choice function.*

It is the planners task to make sure that any admissible population of rational agents represented by some $u \in U$ that obeys the rules designed by him automatically realizes some social state in $F(u)$. To make this idea precise we introduce the concept of a *game form.* A mapping $g : M_1 \times ... \times M_n \longrightarrow A : m := (m_1, ..., m_n) \longmapsto g(m)$ is called an *outcome function.* A tuple $(M_1, ..., M_n, g)$ is called *game form* or *mechanism.* Due to the bijective association between $M := \Pi_{i=1}^n M_i$ and $(M_1, ..., M_n)$ a mechanism is alternatively denoted (M, g). The following observation is fundamental for mechanism theory. For each admissible profile of utility functions $u \in U$ the mechanism (M, g) induces a game $\Gamma_{g,u}$ in strategic form defined by

$$\Gamma_{g,u} := (M_1, ..., M_n; u_1 \circ g, ..., u_n \circ g)$$

.

For obvious reasons $\Gamma_{g,u}$ is also denoted $(M, u \circ g)$.

Denote by $NE(\Gamma_{g,u}), NO(\Gamma_{g,u}) := g(NE(\Gamma_{g,u}))$ and $NP(\Gamma_{g,u}) := u \circ g(NE(\Gamma_{g,u}))$, respectively the sets of Nash equilibria, of Nash equilibrium outcomes and of Nash equilibrium payoffs of $\Gamma_{g,u}$.

Note, that in general the sets $NO(\Gamma_{g,u})$ and $NP(\Gamma_{g,u})$ vary with $u \in U$ as does $F(u)$.

The designer tries to find some mechanism (M, g) such that in any game $\Gamma_{g,u} \equiv (M, u \circ g)$ with $u \in U$ an equilibrium results in a socially desirable outcome. This idea is made precise by the notion of Nash-implementation of a social choice rule.

A mechanism (M, g) *Nash-implements* a social choice rule F on the domain U if $NO(\Gamma_{g,u}) \subset F(u)$ for all $u \in U$.

A large part of the literature uses the concept of **full** implementation requiring equality rather than inclusion. It is full implementation for which Maskin (1999) gave a complete characterization via the properties of *Maskin-monotonicity* and *no veto power*. A careful discussion of the pros and cons of both notions of implementation can be found in Thomson (1996). Jackson (1999) stresses the fact that our (weak) implementation of a social choice rule implies full implementation of some subcorrespondence. We favorize the (weak) implementation, also used by Hurwicz (1994), for the following reasons:

First (weak) Nash implementability does not require Maskin monotonicity. However, the second reason is in fact more important. As long as the social choice rule perfectly describes desirability there is no reason to discriminate between different socially desirable states. Each one is an equally good representative of social desirability. Even if each of the desirable states can be realized by some equilibrium of a game, at most one of these equilibria will be played and thus only one of the desirable states is realized. So performance of the social planners' goal is independent of whether he designs a mechanism that weakly or fully Nash implements the social choice rule.

In the ideal case of implementation by a unique Nash equilibrium the only case without a remaining coordination problem for the planner and the players, only weak Nash implementation is possible. The only exception is a framework where the social choice rule is a social choice *function.* But then weak and full implementation coincide anyway.

There remains only one point to be clarified before we can relate the Nash program to mechanism theory. This is the relation between a solution of a cooperative game say an NTU-game and a social choice rule. While the former one maps cooperative games to sets of feasible payoff vectors for that game the latter one maps profiles of utility functions on outcomes to sets of outcomes. Neither the domains nor the image spaces of these mappings coincide except in very special cases. We shall come back to this problem in the next section.

3 Embedding the Nash program into Mechanism Theory

The literature does not provide much insight into the exact relation between the Nash program and mechanism theory. In Serrano (1997) one finds the statement: *"The Nash program and the abstract theory of implementation are often regarded as unrelated research agendas"*. And Bergin and Duggan (1999) write: *"Nevertheless, because the implementation-theoretic and traditional approaches both involve the construction of games or game forms whose equilibria have specific features, considerable confusion surrounds the relationship between them"*.

In fact there are instances in the literature where the term "implementation" is used in a framework of non-cooperative games where the mechanism theoretic aspect is not addressed at o.k. Sometimes games forms are simply confused with games. Starting from this situation Serrano (1997) attempted *"to clarify the role of the mechanisms used in the Nash program for cooperative games"*. Notice, that even this statement contributes to the confusion by using *"mechanisms"*, a technical synonym for "game form" in the description of non-cooperative foundations that, prior to and without the intended clarification of the relation between the two agendas, cannot justifiably be termed that way. An extension of Serrano's approach is contained in Dagan and Serrano (1998), where in contrast to traditional terminology games in characteristic function form are distinguished from games in coalitional form. The latter are induced by the former ones via adding outcome functions admitting it to define solutions as mappings to outcomes rather than to payoff vectors. These general outcomes extend Serrano's model, where characteristic forms are supplemented by physical allocations resulting from some production economy. The most general model of non-cooperative foundation based on the explicit modelling of physical environments is due to Bergin and Duggan (1999).

Also here cooperative solutions are alternatively defined as mappings resulting in outcomes rather than in payoff vectors. An alternative approach not relying on a specific physical environment is due to Trockel (2000) (see also Naeve (1999)). Here the outcome space needed for an implementation is derived endogeneously from the data of the classes of games considered in the traditional non-cooperative foundation of an axiomatic cooperative solution. We shall illustrate this latter model by the example of the Nash solution in Section 4. A comparison of the different approaches and an evaluation of their respective merits will be part of the concluding Section 5.

Before we can possibly relate the Nash program to mechanism theory it is important to clearly see the differences, the formal ones as well as those in intention and interpretation. Let me recall the latter ones first.

The Nash program relates two alternative ways to model how rational players with partially conflicting goals interact do determine their payoffs. One is a strategic model in which due to lack of commitment power every player non-cooperatively acts on its own. The other one is a coalitional form where for each coalition the feasible payoff vectors are described and axiomatically determined payoff vectors can be contracted on and enforced.

In both approaches the acting persons are players. There is no explicit notion of a society or a social planner or designer. In mechanism theory, in contrast, the only acting person is a designer or social planner. He uses the body of game theory to design general rules forcing any potential population of players from a given pool, by playing equilibria according to those rules, to realize social states he declares *"desirable"* for that population. The players are objects of thought of the designer. So the designer, due to lack of information, enforcement power, and monitoring options, tries to decentralize social decisions uniformly for all feasible populations, trusting to the self-enforcing power of equilibrium.

On the formal level, apart from the players versus designer difference, there are two other crucial differences. One lies on the cooperative side. Following social choice theory mechanism design is based on social choice rules, and therefore, is interested in social states rather than in payoff vectors. Clearly, considering payoff vectors as social states provides one specific degenerate example. The other difference is on the non-cooperative side. To give a non-cooperative procedure for determining social states one has to replace the payoff functions in a game in strategic form by an outcome function associating with any strategy profile some social state.

The link between this outcome based approach and the payoff based Nash program is provided by any population of individuals via their utility functions on the outcome space. These utility functions composed with the outcome function create payoff functions, thereby completing the game form to a game and making the individuals of that population players of this game.

Our next goal is it to illustrate the Nash program, and the problem of Nash implementation by some diagrams. They are modifications of diagrams used by Bergin and Duggan (1999). Furthermore we shall illustrate diagrammatically the difficulties to integrate both diagrams into one and even make this larger one commute.

But beforehand we need some terminology and notation.

Let $\mathcal{G}^C$ and $\mathcal{G}^{NC}$ be sets of cooperative NTU-games in coalitional (or characteristic) form and of non-cooperative games in strategic form, respectively.

By Σ we denote the product of players' strategy sets assumed to be the same for all considered non-cooperative games. This appears not to be a strong restriction, however (cf. Trockel (1999a).

Let A be some non-empty set, interpreted as an outcome space and $h : \Sigma \longrightarrow A$ an outcome function. $\mathcal{U}$ denotes some set of profiles u of utility functions defined on A. L and $\mathcal{L}$ denote a cooperative solution for $\mathcal{G}^C$ and a social choice rule on A, respectively. The use of the symbols "L" and "$\mathcal{L}$" indicates that we finally want the social choice rule $\mathcal{L}$ to suitably represent the solution L.

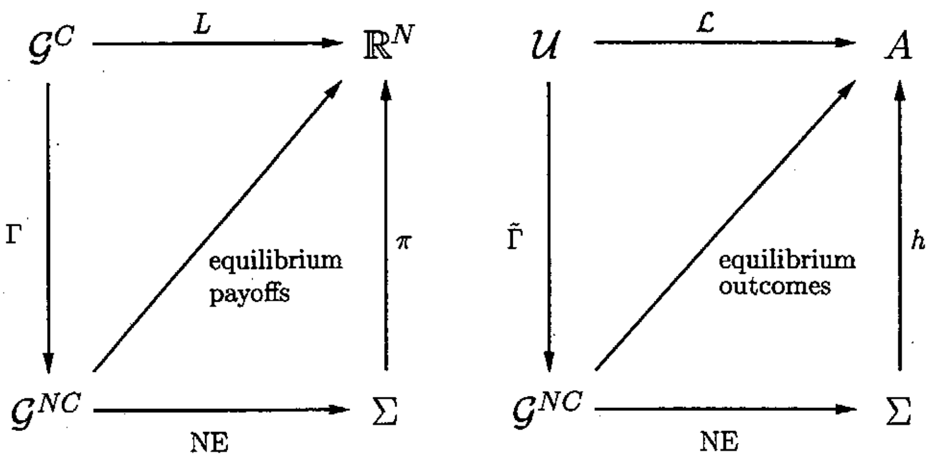

Figure 1:

For any set $\mathcal{F}$ of real-valued functions defined on some non-empty set X the evaluation map, denoted ev, is defined by

$$ev : X \times \mathcal{F} \longrightarrow \mathbb{R} : (x, f) \longmapsto f(x).$$

Figure 1 very clearly shows the differences between the Nash program and implementation theory, despite their obvious structural similarities. A link of the two agendas would require the definition of $\mathcal{L}$ as a suitable representative of L. This in turn requires the specification of a suitable outcome space A. Last not least the relation between $\mathcal{G}^C$ and $\mathcal{U}$ needs clarification.

Figure 2 is a slight modification of Figure 3 in Bergin and Duggan (1999), that represents both situations, the one where one starts with profiles of utility functions and seeks the "right" induced game and the other one, where for a given coalitional form game one looks for profiles of utility functions generating this game. Bergin and Duggan term the

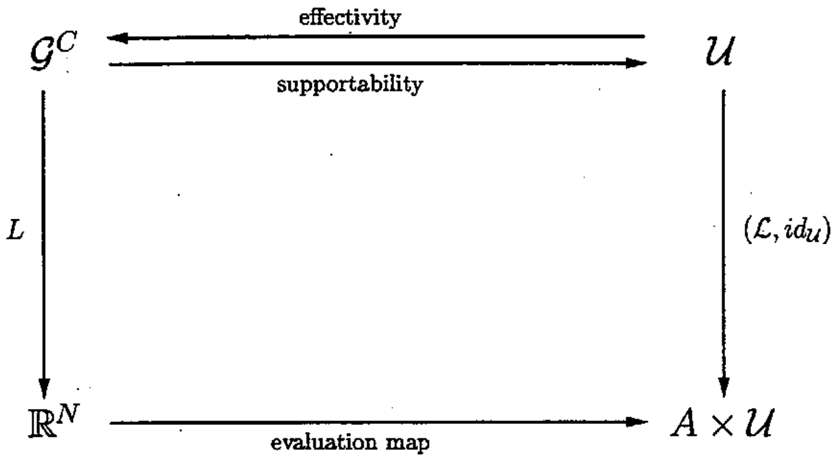

Figure 2:

two problems "effectivity" and "supportability". Clearly, starting with a result from the Nash program and trying to embed this into implementation theory requires to solve the supportability problem. In a situation where both sets $\mathcal{G}^C$ and $\mathcal{U}$ may be identified the supportability and the effectivity problem are solved simultaneously.

The whole problem of embedding the Nash program into mechanism theory is represented by Figures 3 and 4. The inner boldface parts of these diagrams represent the implementation problem, while the outer parts represent the Nash program. The diagrams have to be interpreted as follows:

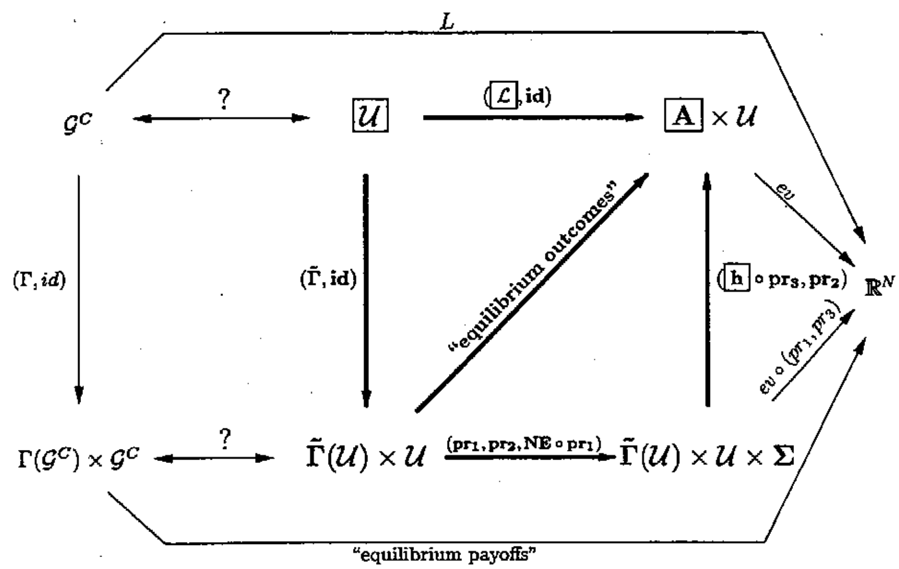

Figure 3:

In mappings between sets we use the arrow "$\longrightarrow$" and do not distinguish between (point-valued) functions and (set-valued) correspondences. In the corresponding diagrams where elements are mapped to elements, respectively, subsets we use "$\longrightarrow$" again. Here the use of "$\in$" versus "$\subset$" indicates whether a point-valued mapping or a correspondence is considered. In Figures 4 and 5 the question marks and the boxes marking some of the sets indicate the explananda. The embedding of the Nash program into mechanism theory is only possible if the explananda can be consistently defined such that the diagrams commute.

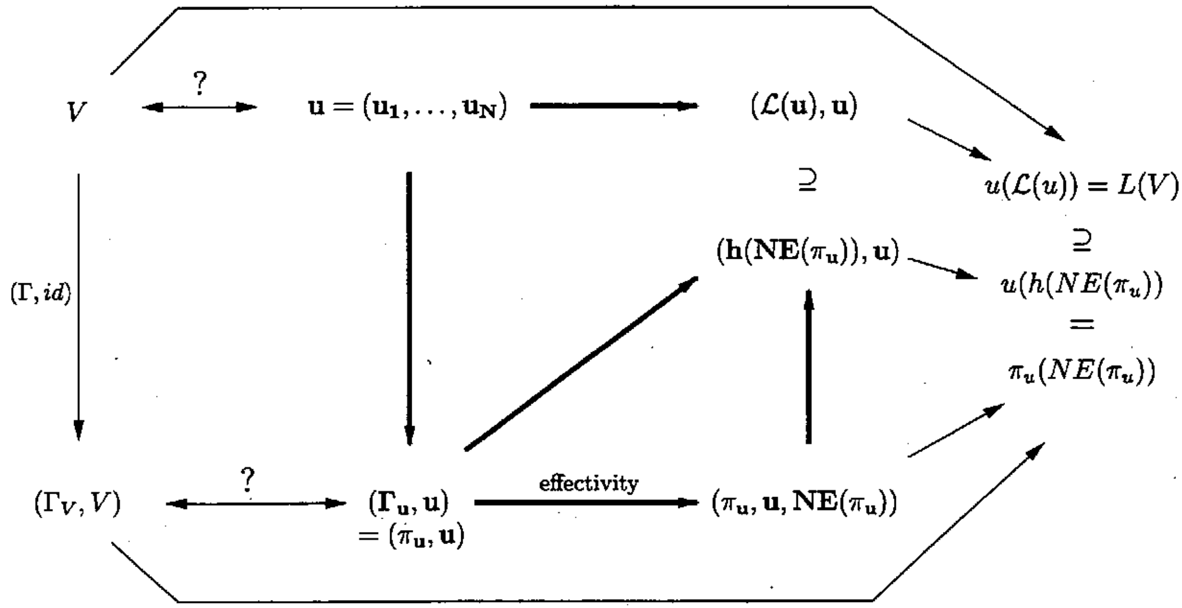

Figure 4:

That this problem can be solved has been proved by Trockel (1999a) His *Embedding Principle* for the Nash program may be stated as follows:

Assume for any game $V \subset \mathcal{G}^C$ and its associated game $\Gamma^V \in \mathcal{G}^{NC}$ one knows that $\pi^V(\Sigma)$ is feasible for V and $\pi^V(NE(\Gamma^V)) \subset L(V) \neq \emptyset$. Then a mechanism (Σ, h) exists that weakly Nash-implements $\mathcal{L}$. Moreover for every $V \in \mathcal{G}^C$ one has: $u^V \circ \mathcal{L}(V) = L(V)$.

Clearly, to prove this statement definitions of the outcome function h and of the social choice rule $\mathcal{L}$ are required. Also a specification of how $\mathcal{G}^C$ and $\mathcal{U}$ are defined and identified is needed. While these details may be found in Trockel (1999a) we shall sketch the proof of the Embedding Principle for the special class of two-person bargaining games and the Nash solution in Section 4.

4 An Example: The Nash Bargaining Solution

4.1 Impossibility of Nash-implementation

The claim in the literature that the Nash bargaining solution cannot be Nash implemented is based on the proof that it is not Maskin monotonic. As Maskin monotonicity is a necessary assumption for full Nash implementability of social choice rules this claim relies on some implicit assumptions. First, it concerns only full in contrast to weak implementation. Secondly, it takes the interpretability of the Nash solution as a social choice rule for granted. The outcome space for which the lack of Maskin monotonicity is demonstrated is, however, not arbitrary, it is rather a set of lotteries over some finite set. The essence of the argument can be seen in the following example which is a simplification of an example due to Howard (1992).

Let $A = \{a, b, c\}$ an outcome space. Let $\{u, v\}$ be the set of admissible profiles of utility functions on A, with $u = (u_1, u_2), v = (v_1, v_2)$ defined by:

$$u_1(a) = 0 = u_2(a), u_1(b) = 1/8, u_2(b) = 1, u_1(c) = 1/2 = u_2(c)$$

$$v_1(a) = 0 = v_2(a), v_1(b) = 1/8, v_2(b) = 1, v_1(c) = 1/2 = v_2(c).$$

The Nash social choice rule is defined by

$$\mathcal{N} : \{u, v\} \Longrightarrow A : w \longmapsto \mathcal{N}w = argmax_{x \in A}\ w_1(x)w_2(x).$$

Notice that even in this most simple framework $\mathcal{N}$ *does not* coincide with the Nash solution N. Rather N is defined by

$$N : \{u(A), v(A)\} \Longrightarrow [0,1]^2 : w(A) \longmapsto N(w(A)) = argmax_{y \in w(A)}\ y_1 y_2.$$

We get $\mathcal{N}(u) = \{c\}$ and $\mathcal{N}(v) = \{b\}$. Going from u to v does not involve a preference reversal of any member ot the society. Nevertheless, the socially desired state does change! So the social choice rule fails to be Maskin monotonic. Hence, *in this context* the Nash solution represented by $\mathcal{N}$ is not Nash-implementable.

As we shall show the choice of the outcome space is crucial for the problem of Nash implementability of the Nash solution.

4.2 A Support Result

Consider a two person bargaining situation S as illustrated in Figure 6.

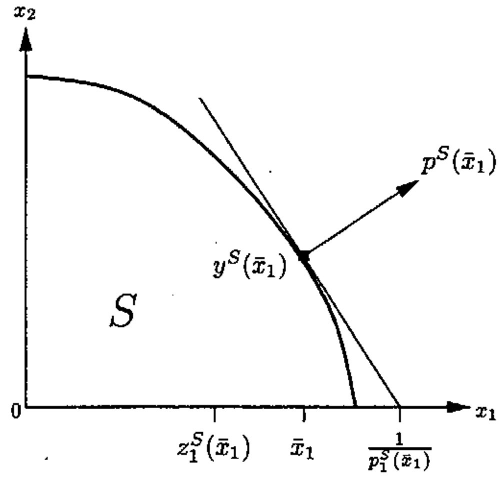

Figure 5:

The compact strictly convex set $S \subset \mathbb{R}^2$ represents all feasible utility allocations for two players. For simplicity assume that the efficient boundary ∂S of S is the graph of some smooth decreasing concave function from $[0, 1]$ to $[0, 1]$. Such a bargaining situation can be looked at as a two-person NTU-game, where S is the set of payoff vectors feasible for the grand coalition $\{1, 2\}$, while $\{0\}$ represents the payoffs for the one player coalitions. The normalization to $(0; 1, 1)$ is standard and reflects the idea that S arose as the image under the two players' cardinal utility functions of some underlying set of outcomes or allocations. Cardinality determines utility functions only up to positive affine transformations and therefore justifies our normalization. Now, consider the following modification of Nash's simple demand game due to Trockel (2000)

$$\Gamma^S = (\Sigma_1, \Sigma_2; \pi_1^S, \pi_2^S).$$

$\Sigma_1 = \Sigma_2 = [0, 1]$ are the players' sets of (pure) strategies. The payoff functions are defined by $\pi_i^S(x_1, x_2) := x_i \mathbf{1}_S(x_1, x_2) + z_i^S(x_i)\mathbf{1}_{S^C}(x_1, x_2)$.

Here S^C is the complement of S in $[0, 1]^2$ and $\mathbf{1}_S$ is the indicator function for the set S. Finally $z_i^S(x_i)$ is defined as follows:

For each $x_i \in [0, 1]$ the point $y^S(x_i)$ is the unique point on ∂S with $y_i^S(x_i) = x_i$. By $p^S(x_i)$ we denote the normal vector to ∂S at $y^S(x_i)$ normalized by $p^S(x_i) \cdot y^S(x_i) = 1$. Now $z_i^S(x_i)$ is defined by $z_i^S(x_i) = min(x_i, \frac{1}{2p_i^S(x_i)})$, $i = 1, 2$.

This game has a unique Nash equilibrium (x_1^*, x_2^*) that is strict, has the maxmin-property and coincides with the Nash solution of S, i.e. $\{(x_1^*, x_2^*)\} = N(S)$. The idea behind the payoff functions is it to consider for any efficient utility allocation y its value under the efficiency price vector $p(y)$. If the utility allocation could be sold at $p(y)$ on a hypothetical market and the revenue would be split equally among the players there is only **one** utility allocation such that both players could buy back their own utility with their incomes without the need of any transfer of revenue. This equal split of revenue in the payoff function corresponds to *equity* in Shapley's (1969) cooperative characterization of the λ-transfer value via equity and efficiency. As for our two-person bargaining games the λ-transfer value just singles out the Nash solution this result does not come as a big surprise. By supplementing efficiency, which characterizes the infinitely many equilibria in Nash's demand game, by the additional equity, embodied in the payoff functions $\pi_i^S, i = 1, 2$, one gets the Nash solution as the unique equilibrium of the modified demand game. This result provides obviously a non-cooperative foundation of the Nash solution in the sense of the Nash program. If the impossibility of Nash implementation of the Nash solution as claimed in the literature [Howard(1992), Serrano(1997), Dagan and Serrano (1998)] would hold generally true, independently of the choice of the outcome space and the resulting specification of the Nash social choice rule, then it would be impossible to extend our above foundation to an implementation result. As we shall see, however, this is not the case. We shall provide a framework in which the Nash solution *can* be Nash implemented.

4.3 Nash-implementation

According to the diagrams in Figures 3 and 4 to extend our non-cooperative foundation to Nash implementation we have to perform the following tasks:

1. Define the outcome space A.
2. Specify the set $\mathcal{U}$ of feasible profiles of utility functions on A.
3. Represent the Nash solution $L := N$ by a social choice rule $\mathcal{L} := \mathcal{N}$.
4. Clarify the relation between $\mathcal{G}^C$ and $\mathcal{U}$.
5. Define the outcome function h.

The set $\mathcal{G}^C$ is in our present context the set of all $(0; 1, 1)$-normalized two-person bargaining games, $V = S$, as defined above. Let us now define the outcome space A by

$$A := \{L \in ([0,1]^2)^{\mathcal{G}^C} | \forall S \in \mathcal{G}^C : L(S) \in S\}.$$

From now on we look at the singleton-valued correspondences L as functions.

So A is the set of all possible bargaining solutions on $\mathcal{G}^C$.

Next, we perform our tasks 2 and 4.

Every bargaining game $S \in \mathcal{G}^C$ induces a profile of utility functions $u^S := (u_1^S, u_2^S)$ on A as follows:

$$u_i^S(L) := L(S)_i \equiv: L_i(S), i = 1, 2.$$

The interpretation is that player i in the game S has a utility function on the set of bargaining solutions dependent only on the game S he is involved in. The more a solution pays him out in that game the higher is his utility of that solution.

For each player $i \in \{1, 2\}$ different games S define different utility functions u_i^S. Hence, the map $S \longmapsto u^S = (u_1^S, u_2^S)$ is an injection that allows us to identify the set $\mathcal{G}^C$ with a set of utility profiles. In fact, we define $\mathcal{U}$ as the image under that map of $\mathcal{G}^C$ and thereby solve for our purpose the effectivity and supportability problems illustrated in Figure 3.

The next task left is the third one, i.e. the definition of the Nash social choice rule $\mathcal{N}$. To accomplish this we define for any $S \in \mathcal{G}^C$ an equivalence relation $\sim_S$ on the set A of bargaining solutions by

$$L \sim_S L' :\Longleftrightarrow L(S) = L(S').$$

The set $[L]_S := \{L' \in A | L' \sim_S L\}$ is the S-equivalence class generated by the solution L.

Now the Nash social choice rule is the correspondence

$$\mathcal{N} : \mathcal{G}^C \Longrightarrow A : S \longmapsto [N]_S \subset A$$

.

Notice, that due to the identification of $\mathcal{U}$ and $\mathcal{G}^C$ this correspondence *is* really a social choice rule.

Finally, we come to our fifth and last task, which is to define the game form $([0,1]^2, h)$. The outcome function $h : [0,1]^2 \longrightarrow A : x = (x_1, x_2) \longmapsto h(x)$ maps every possible strategy profile to some bargaining solution. Notice, that this, as a mapping from $\mathcal{G}^C$ to $[0,1]^2$ has to be defined by specifying pointwise, i.e. for every $S \in \mathcal{G}^C$, the element $h(x)(S) \in S \subset [0,1]^2$.

Accordingly, given any strategy profile $x \in [0,1]^2$, we define for any $S \in \mathcal{G}^C$ an element of S by $h(x)(S) := \pi^S(x)$. The $\pi^S = (\pi_1^S, \pi_2^S)$ is the profile of payoff functions of our demand game defined above. So the social state $h(x)$ associated by the outcome function h with the strategy profile x is the mapping $S \longmapsto \pi^S(x)$ defined on $\mathcal{G}^C$.

It is very important to notice that a social planner may very well know that mapping without having any knowledge about the utilities of any two players meeting in a specific bargaining game S. Put differently, for any $x \in [0,1]^2$ the *outcome* $h(x) = \pi^{(\cdot)}(x)$ is defined independently of any specific game $S \in \mathcal{G}^C$. Clearly, the *images* of different games $S \in \mathcal{G}^C$ vary depending on S. Now it is not hard to see that $([0,1]^2, h)$ indeed Nash-implements the social choice rule $\mathcal{N}$. For the game $\Gamma^S := (\Sigma_1, \Sigma_2; u_1^S \circ h, u_2^S \circ h)$, with $\Sigma_1 = \Sigma_2 = [0,1]$ we just have to insert the point $N(S)$ for x in $h(x)(S)$ to get $h(N(S))(S) = \pi^S(N(S)) = N(S)$.

This equality asserts that the solution $h(N(S))$ is S-equivalent to the Nash solution N, i.e. $h(N(S)) \in [N]_S \equiv \mathcal{N}(S)$. As for the unique Nash equilibrium $x^{*S} = (x_1^{*S}, x_2^{*S})$ of Γ^S we had $NE(\Gamma^S) \equiv \{x^{*S}\} = \{N(S)\}$ we get $h(NE(\Gamma^S)) = h\{N(S)\} = \{h(N(S))\} \subset [N]_S \equiv \mathcal{N}(S)$.

This establishes (weak) Nash implementation of the Nash social choice rule N. Notice, that $\mathcal{N}$ and N, though intimately related, are different mathematical objects. Nevertheless the term *"Nash implementation of the Nash solution"* is justified as uniformly for all games $S \in \mathcal{G}^C$ the payoffs to the players according to the unique Nash equilibrium are identical to the payoffs defined by the Nash solution. This follows immediately from the following chain of equalities:

$$u_i^S \circ h(N(S)) = u_i^S(h(N(S))) = u_i^S(N) = N_i(S), i = 1, 2.$$

So we ended up with weak implementation in unique Nash equilibrium. For practical reasons this is as good as would have been a weak implementation of the constant social choice rule $\tilde{N} : \mathcal{U} = \mathcal{G}^C \longrightarrow A$ defined by $\tilde{N}(u) := N$, which in fact would be even full implementation. Notice, that whatever S from the pool $\mathcal{G}^C$ is going to materialize, the induced game Γ^S by its unique equilibrium exactly determines the Nash solution payoffs $N(S)$. From the point of view of observability it is impossible, hence even meaningless to tell whether $\tilde{N}$ or N has materialized. In a welfaristic framework, where payoffs rather than social states are the objects of interest, we just do not care.

After we have established weak Nash implementability of $\mathcal{N}$ its Maskin monotonicity is not really an issue. Nevertheless it is interesting to see the impact the framework, in particular the choice of the outcome space, has on this matter. In fact it turns out that our Nash social choice rule $\mathcal{N}$ is Maskin monotonic. To see this we have to demonstrate that any change of preference profiles that results in a socially desired state according to $\mathcal{N}$ not socially desired at the original preference profile must necessarily involve a preference reversal.

So start with a preference profile $(\succsim_1^S, \succsim_2^S)$ on A represented by (u_1^S, u_2^S) and consider a second profile $(\succsim_1^{S'}, \succsim_2^{S'})$ represented by $(u_1^{S'}, u_2^{S'})$.

Now, let L' be socially desired for S' but not for S, i.e. $L'(S') = u^{S'}(L') = u^{S'}(N) = N(S')$ but $L'(S) = u^S(L') \neq u^S(N) = N(S)$. So we have $u_i^{S'}(L') \geqslant u_i^{S'}(N), i = 1, 2$.

But as $u^S(N)$ is Pareto-efficient in S but different from $u^S(L')$, for at least one player $j \in \{1, 2\}$ we get $u_j^S(N) > u_j^S(L')$. This establishes the preference reversal. Hence, $\mathcal{N}$ is

Maskin monotonic.

It is easy to see that in this framework any Pareto efficient social choice rule, as for instance $\tilde{N}$ is Maskin monotonic, too (cf. Trockel (1999a). An interesting question that could be asked now is as to full implementation of the social choice correspondence $\mathcal{N}$.

Clearly, this can be possibly achieved only if the strategic games employed have multiple equilibria. In fact, a game doing this job can be easily generated from our game Γ^S. It has been used in Trockel (2001) to provide for the class of all bargaining solutions a meta bargaining result akin to the one of van Damme (1986) and its extension due to Naeve-Steinweg (1999) who restrict to bargaining solutions satisfying certain plausible axioms.

The construction is quite simple. Define the game $\bar{\Gamma}^S$, a *"meta-bargaining game"* by $\bar{\Gamma}_S = (A, A; \bar{\pi}_1^S, \bar{\pi}_2^S)$.

As A is the set of all bargaining solutions it remains to define $\bar{\pi}_i^S, i = 1, 2$. To do this consider for any $S \in \mathcal{G}^C$ the mapping $E_S := proj_{1,4} \circ ev_S : A \times A \longrightarrow [0,1] \times [0,1] : (L^1, L^2) \longmapsto (L_1^1(S), L_2^2(S))$. Now we define $\bar{\pi}_i^S : A \longrightarrow [0,1]$ by $\bar{\pi}_i^S = \pi_i^S \circ E_S, i = 1, 2$.

In this game any pair of solutions $(L^1, L^2) \in [N]_S \times [N]_S$ is a Nash equilibrium. The only pair of bargaining solutions that is an equilibrium for each $S \in \mathcal{G}^C$ is the pair (N, N).

By a similar procedure as explained above for Γ^S also $\bar{\Gamma}^S$ may be used to derive Nash implementation of $\mathcal{N}$. But now, this is full implementation. Hence the Maskin monotonicity of $\mathcal{N}$, which we have already established, is a necessary consequence.

On the first view we seem to be in conflict now with Result 2 in Dagan and Serrano (1998). According to their result our $\mathcal{N}$ should be ordinally invariant. That means that for any monotonic transformation $T : \mathbb{R}_+^2 \longrightarrow \mathbb{R}_+^2$ such that $T(S) \in \mathcal{G}^C$ we should have $\mathcal{N}(S) = \mathcal{N}(T(S))$. But, as is well known for such a transformation and the Nash solution N one has in general $T(N(S)) \neq N(T(S))$. Now consider a solution L such that $L(T(S)) = T(N(S))$ and $L(S) = N(S)$. Then $L \in [N]_S = \mathcal{N}(S)$, but $L \notin [N]_{T(S)} = \mathcal{N}(T(S))$. Hence $\mathcal{N}(S) \neq \mathcal{N}(T(S))$.

The source for the seeming contradiction lies in the restrictive notion of implementation used by Dagan and Serrano (1998) which explicitly makes use of a set of probability distributions as an outcome space. One can read in their Introduction *"that if a solution concept is independent of randomization it must be ordinally invariant... This implies that major solution concepts in coalitional games (eg. the Nash bargaining solution...) can be derived strategically only by considering the possibility of random outcomes: either chance moves, mixed strategies, or pure strategy equilibrium refinements based on trembles must be past of the analysis."*

Our approach is a counter example to this general claim. Again this is due to the fact that their results are derived in a special model with a specific notion of implementation and a specific outcome space.

5 Conclusions

5.1 Answers to the Questions

What we have demonstrated in Section 4 via the example of the Nash bargaining solution, namely, how to get an implementation from simple non-cooperative foundation, can be performed in a very general way for any solution concept on any class of coalitional games. The trick is that the outcome space has to be the set of singleton-valued solutions (cf. Trockel (1999a)).

So formally the answer to the first question in the title is "yes". The Nash program can be seen as part of mechanism theory. One needs a quite complex outcome space and one needs a restricted domain of preference profiles on that space. Also the set of preference profiles is not a product set. The preferences in a profile are dependent on each other. But this is unavoidable in a welfaristic framework. Take for instance a bargaining game S. What remains of this S if by going from games to game forms information on players utilities is removed? As all that information is contained in S or, in its efficient boundary there just remains the set $[0,1]^2$, the smallest set product containing all $S \in \mathcal{G}^C$ and taken as strategy space in our approach. The only way then to speak of say, the Nash solution is not by locating it for a given S. There is no such *given* S anymore! Rather we may speak of the Nash solution as of an abstract concept representing some possible social rule for how to solve bargaining problems. Our approach makes any solution a *social state.*

This specific choice of an outcome space is perfectly in accordance with the general framework of implementation theory and mechanism design. Formally, it is as adequate as some set of lotteries, some space of allocations of commodities, or some set of candidates in a context of voting. Only, the latter ones reflect special economic or social scenarios that are modelled in a game theoretic language. Then the respective physical outcome space becomes *the outcome space* in the accordingly modelled mechanism.

But, once in the welfaristic purely game theoretic world, notions like candidate or allocations do not have a place. At best they help to clarify the formal model as intuitive examples for applications.

The space of solutions for a considered class of games, in contrast, is endogeneously determined rather than derived from any specific application. So, formally, we have a purely game theoretic model by which we may extend non-cooperative foundations of cooperative solutions to Nash implementations. This does not say that this is of any use for applications. So the second question is much harder to answer. *Should we see the Nash program as a part of mechanism theory?* The answer clearly depends on the task one has in mind. I had elaborated already on the Nash program and its supposed value and intuition. One may wonder what the possibility of Nash implementability might add to these considerations. What gain does it bring to the Nash program? Alternatively, one might ask, like Serrano (1997), whether using the Nash program brings any advantage for mechanism theory.

Let me deal with the first question first. Assume two players are in a concrete bargaining situation, which they both agreed to be suitably modelled by some bargaining game $S \in \mathcal{G}^C$. Among the two there is common knowledge about this S. Now suppose during the process of looking for an acceptable point in S somebody offers the two players to solve their problem by committing to play some game in strategic form instead. And suppose they would agree to do so. Would the fact that this game is constructed from their specific S and some game form *independent* of S be of any help for them? Probably not. They would not care whether their payoff functions could be factorized in such a way. They just would look at the possible payoff vectors of the game, in equilibrium and outside. They might reject a game that modifies Nash's demand game in such a way that each player receives his coordinate of N if he chooses it as a strategy and gets 0 otherwise. Accepting this game would be a tatonnement to finishing their negotiation process by agreeing on N.

Assume they are offered instead to play Γ^S as defined above. Now all feasible points in S still remain accessible via coordinates strategic actions, in particular the Nash solution. But in case of disagreement by playing the equilibrium they would get N rather than falling back to 0. So the commitment would be a Pareto improvement in case of disagreement.

It would be advisable for both players to accept to play that game. But this fact is independent of the mechanism theoretic implementability. So I do not see that the possibility of extending a non-cooperative foundation to implementation is of any value for the players looked at in the Nash program. It is, however, of potential value for the game theorist. The ways how payoff functions may be possibly factorized into outcome function and utility functions as explained by implementation may help to distinguish between more or less "sensible" strategic models for non-cooperative foundation. What does mechanism theory gain from the embedding possibility? One might follow Serrano (1997) by arguing that the Nash program utilized by the embedding principle provides a useful pool of potential mechanism like the revelation principle. But this answer is too general. If a planner or designer is in the situation of looking for the possibility to implement a social choice rule its outcome space is usually given as are institutional restrictions on the potential rules for the game form. This would require more "concrete" outcome spaces and therefore be in conflict with the welfaristic framework of the Nash program.

Clearly, even a very abstract outcome space like a set of solutions that we used above is *in principle* usable. It only puts extreme requirements in terms of observability and enforcement tools upon the planner. It is not so much the problem that the planner has to check whether the (an) equilibrium has been played. This is superfluous due to the self-enforcing power we use to desribe to equilibrium. If this could and would have to be checked there would be no need in insisting on equilibrium play. Any play could be checked and therefore might be prescribed by the planner. It is rather the problem to make sure that the players really *play* the game they are supposed to. It might not be in their interest to play the game. An indirect control would be possible in some cases via observation of outcomes. Here more concrete physical outcomes may be much easier to be monitored from outcomes that are abstract principles, norms or solutions. The solution might be observable via the payoff it prescribes in a game. But this is possible only if the game is *known* to the observer, which implies that players' preferences have to be known. But then implementation via game forms is an unnecessary complication.

Hence my conclusion is that mechanism theory does not profit very much, either, from its formal access to the Nash program. So perhaps, we should not look at the Nash program from this point of view.

5.2 Outlook

What are the consequences from our considerations? We should proceed beyond present mechanism theory. We need a theory that tells us which are "good" or "bad" mechanisms and this qualification should be dependent on the information and enforcement possibilities that are available in the scenario under consideration. This leads to *genuine implementation* as advocated by Hurwicz (1994). It requires treatment of institutional features as part of the implementation problem. One also needs to explicitly address the problem of voluntary implementation like in Jackson and Palfrey (2001).

Up to now this aspect had been treated only in a stylized rudimentary form as participation constraint in agency theory.

A further research agenda suggested by our above considerations is a kind of a generalized Nash program consisting of several parts. One part would stress the symmetry of the cooperative and the non-cooperative approach in foundations of game theory and accordingly look also for cooperative foundations of non-cooperative equilibria. A second part would extend the applications from linking just game theoretic cooperative solutions and non-cooperative equilibria to any kind of social or economic solution concepts, like Walrasian equilibria, evolutionary stable or institutionally distinguished outcomes. Further, extensions to non-welfaristic social scenarios should be considered. See for instance, Sertel and Yildiz (2001) and Sotskov (2001).

References

[1] Bergin, J., Duggan, J.:"Implementation-theoretic approach to non-cooperative foundations", *Journal of Economic Theory* **86**, 50–76 (1999)

[2] Binmore, K.: "Nash bargaining theory I, II" in: Binmore, K., Dasgupta, P. (eds.): The economics of bargaining, Cambridge: Basic Blackwell 1987

[3] Binmore, K.:"Introduction" in: Nash, J.F. Jr.(ed.): Essays on game theory, Cheltenham: Edward Elgar 1997

[4] Binmore, K., Dasgupta, P.:"Introduction" in: Binmore K. and Dasgupta, P.(eds.): The Economics of Bargaining, Cambridge, Basis Blackwell 1987

[5] Binmore, K., Rubinstein, A., Wolinsky, A.: "The Nash bargaining solution in economic modelling", *Rand Journal of Economics*, **17**, 176–188 (1986)

[6] Crawford, V.P.: "A procedure for generating Pareto-efficient egalitarian equivalent allocations", *Econometrica*, **47**, 49–60 (1978)

[7] Dagan, N., Serrano, R.:"Invariance and randomness in the Nash program for coalitional games", Economics Letters **58**, 43–49 (1998)

[8] Haake, C.-J.: "Supporting the Kalai-Smorodinsky bargaining solution", Bielefeld University, IMW Working Paper No. 301 (revised version) (2000)

[9] Howard, J.V.: "A social choice rule and its implementation in perfect equilibrium", *Journal of Economic Theory* **56**, 142–159 (1992)

[10] Hurwicz, L.: "Economic design, adjustment processes, mechanisms and institutions", *Economic Design* **1**, 1–14 (1994)

[11] Jackson, M.O..: "A crash course in implementation theory", forthcoming, Thomson, W.(ed.): The axiomatic method: "Principles and Applications to game theory and resource allocation", 2001

[12] Jackson, M.O., Palfrey, T: "voluntary implementation", *Journal of Economic Theory,* forthcoming, 2001

[13] Maskin, E.S.: "Nash equilibrium and welfare optimality", *The Review of Economic Studies* **66**, 23–38 (1999)

[14] Moulin, H.: "Implementing the Kalai-Smorodinsky bargaining solution", *Journal of Economic Theory* **33**, 32–45 (1984)

[15] Nash, J.F.:"Non-cooperative games", *Annals of Mathematica* **54(2)**, 286–295 (1951)

[16] Nash, J.F.:"Two-person cooperative games", *Econometrica* **21**, 128–140 (1953)

[17] Naeve, J.: "Nash implementation of the Nash bargaining solution using intuitive message spaces",*Economics Letters* **62**, 23–28 (1999)

[18] Naeve–Steinweg, E.: "A note on van Damme's mechanism" *Review of Economic Design* **4**, 179–187 (1999)

[19] Osborne, M.J., Rubinstein, A.: "Bargaining and markets", New York: Academic Press 1990

[20] Rubinstein, A.: "Perfect equilibrium in a bargaining model", *Econometrica* **50**, 97–109 (1982)

[21] Selten, R.: "Re-examination of the perfectness concept for equilibrium points in extensive games" *International Journal of Game Theory* **4**, 25–55 (1975)

[22] Selten, R.: "Oral communication" (1999)

[23] Serrano, R.: "A comment on the Nash program and the theory of implementation", *Economics Letters* **55**, 203-208 (1997).

[24] Sertel, M.R., Yildiz, M.: "The impossibility of Walrasian bargaining solution", forthcoming in: Sertel, M.R. (ed.): Advances in economic design, Springer, Heidelberg, 2001

[25] Sotskov, V.:"Characterization of competitive allocations and the Nash bargaining problem, forthcoming in: Sertel, M.R. (ed.): Advances in economic design, Springer, Heidelberg, 2001

[26] Trockel, W.: "A Walrasian approach to bargaining games", *Economics Letters* **51**, 295–301 (1996)

[27] Trockel, W.: "Integrating the Nash program into mechanism design", UCLA Working Paper No. 787 (1999a)

[28] Trockel, W.: "Unique implementation for a class of bargaining solutions", *International Game Theory Review* **1**, 267–272 (1999b)

[29] Trockel, W.: "Implementation of the Nash solution based on its Walrasian characterization" *Economic Theory* **16**, 277–294 (2000)

[30] Trockel, W.: "A universal meta bargaining realization of the Nash solution", *Social Choice and Welfare*, forthcoming (2001)

[31] Van Damme, E.: "The Nash bargaining solution is optimal", *Journal of Economic Theory* **38**, 78–100 (1986)