Working Paper 9016

SOME PROBLEMS OF INFINITE REGRESS IN SOCIAL-CHOICE MODELS: A CATEGORY THEORY SOLUTION

by Fadi Alameddine

Fadi Alameddine is a research assistant at the Federal Reserve Bank of Cleveland. The author wishes to thank Paul Bauer, Randall Eberts, Owen Humpage, Robin Ratliff, and especially Spyros Vassilakis, whose research and comments have been instrumental to this study.

December 1990

Introduction

In modern Western democracies, economic and political institutions often have been criticized on moral grounds. The arguments pinpoint the resulting inequalities and inefficiencies as the evidence of these institutions' inadequacy to provide justice. However, evaluating institutions in retrospect (ex post), by contrasting their ex-post resource allocation with other allocations known to be feasible ex post, is misleading. Social decisions must be made under conditions of uncertainty. Hence, institutions must be evaluated before the uncertainty is resolved (ex ante), i.e., according to their expected performance, as delimited by the information available at the time decisions are made. So an institution can be condemned only if an alternative one exists yielding preferable outcomes (by any measure to be decided upon) under the same ex-ante information set.

This view is mainly endorsed by economists in the latter part of the twentieth century. However, it ignores the process by which a given institution is to be chosen. The process is itself an institution, so that the statement of the problem embeds in an infinite regress: The process must itself be the object of choice of some process which, once again, is itself an institution and therefore the object of choice of another process. Unless this infinite regress is resolved, further inquiries on social-choice theory will be limited in scope. Furthermore, because institutional decisions are actually made, a proper theoretical account must capture the process.

This paper is concerned with the solution of the infinite regress problem as it arises in social-choice theory. My approach is to draw on arguments that consider social choices in game-theoretic environments where information is decentralized, and whose fundamental outcomes are the procedures conducive for players to coordinate decisions. In the tradition of game theory, these procedures are called *mechanisms*. The objective is to establish the existence of a *universal* mechanism, i.e., a mechanism whose strategy spaces include all possible proposals to change the mechanism.

Section I outlines what I have termed the "Gauthier framework" and elaborates how the infinite regress arises. Section II maps out a general style for solving recursive specifications, with category theory providing the concepts needed for the construction of a universal mechanism. Section III adapts the universal mechanism construction, in Vassilakis (1989), to the Nash demand game, which can be seen as a formal Gauthier bargaining setup.

I. The Infinite Regress Statement Within Gauthier's Framework

The analogy between individual choice and social choice has been aptly formulated by John Rawls (1971): As each individual rationally decides what constitutes his own good, a society must decide on its system of justice. Thus, a society must be modeled as if it has an objective function. In decision theory, the view supporting rational individual choice is that of expected utility maximization. The task of social-choice theory is to construct an objective function from individual utility functions. Social choice then must be a product of meditated choice by its individuals.

To extract a social-objective function (social-welfare function), Rawls suggests putting any rational individual behind the veil of ignorance and asking that person to select the basic rules for a society. This person's chosen rules are the principles of justice, to which all persons should agree. As David Gauthier remarked, Rawls' approach is not a viable solution for the rational-choice problem because it requires individuals to form a concept of justice prior to the original agreement. From the parallel between the individual and social choice, such a prior concept cannot be justified. It is itself a principle of justice and should be the outcome rather than the assumption of the theory.

The individual preferences (utilities) must be the only argument of the social-welfare function. If society's preference is represented by a mapping, then the question becomes how should it depend on individual utility functions?

My stand echoes Gauthier's (1984, p. 255): "...the principles of justice are those principles for making social decisions or choices to which rational individuals, each seeking to cooperate with her fellows in order to maximize her own utility, would agree." A desirable social welfare function is not the outcome of single optimization problems, but the outcome of mutually consistent optimization problems. The mutual consistency is formalized by the solution concept of a game.

Gauthier proceeds to propose a bargaining game in the spirit of a Nash demand game.[1] However, the game's bargaining procedures are exogenously specified (see Gauthier [1982], p. 256). For example, he insists that all parties must be equally able to advocate and advance their interests (fair play). Furthermore, all agents must have an identical information set when decisions are made. He then shows that all players have the same dominant strategy, whose outcome is *individually rational* (weakly preferred by all agents over nonparticipation), and *incentive compatible* (truthful reporting is a dominant strategy for all players).[2]

There are two ways to build on Gauthier's work. The first is to relax the complete information assumption. In bargaining games, allocations are type contingent (an agent's type is usually characterized by his preferences and information set). In general, agents will not take the same actions in

* * * * * * * * * * * * *

1 Length constraints preclude a presentation of a formal account of Gauthier's game. Please see the referenced work for a full exposition.

2 Gauthier does not explicitly use these terms. However, by a Coasian argument, these results can be directly extracted.

equilibrium because their actions reveal their private information, which might have an adverse effect on their payoff. This, in turn, restricts the set of feasible mechanisms. To illustrate, the appendix contains a simple voting game, first advanced by Holmstrom and Myerson (1983).

Second, the comments directed at Rawls' veil of ignorance can be channeled against any exogenous condition that the philosopher imposes. Fair play (on the procedural level) is itself a principle of justice and hence should not be assumed. One could surmise that fair play must be agreed upon by some prior bargaining session, but the rules of this session cannot be exogenously imposed either. Therefore, we must construct yet another bargaining mechanism to solve the problem at this level.[3] Clearly, this problem will appear at every level. The resulting infinite regress must be dealt with by constructing a mechanism in which strategy spaces contain proposals for amending the rules of any game; I call this mechanism *universal.*

This argument is beyond a mere technicality. If we are to extract the principles of justice from the equilibrium of a game based on a given mechanism, then it is imperative to show that such a mechanism is not only feasible, but that it is chosen exclusively by the players, and not introduced exogenously by the analyst through the procedural rules. Consider a game, representing a society, with n agents faced with the problem of allocating the resources in their economy. I refer to this game as the *underlying game.* Initially, all agents have a given endowment and a set of strategies. If play

* * * * * * * * * * * *

3 This idea is advanced in Crawford (1985).

takes place noncooperatively, i.e., a "war of all against all" scenario, the resulting Nash equilibrium is usually Pareto inefficient. As a parable, we can think of the well-known prisoners' dilemma game. The Nash equilibrium coincides with both players confessing, which is clearly suboptimal to both not confessing.[4] Now, agents can suggest Pareto improvements by attempting *preplay negotiations*. Players try to coordinate their actions through mechanisms that determine the play of the underlying game. Preplay negotiations can be viewed as a separate round of bargaining over mechanisms, but there is no guarantee that an agreement will be reached in this game either. The point is that whatever procedural constraints the philosopher wishes to introduce, he must show them to be the outcome of a previous stage of bargaining. Furthermore, this point remains valid whether the constraints are introduced on the procedures of the underlying game or on the procedures of the subsequent preplay negotiations.

This section concludes with a brief discussion concerning the solution concept of a game and Nash equilibria. Although it may appear that this paper has identified a Nash equilibrium as the solution to a game, this view is misleading. Instead, I present the concept of a Nash equilibrium along the same lines as Kreps (1990): "The concept (of a Nash equilibrium) is advanced as an answer to the question: If there is an obvious way to play the game (a way that all players can figure out and all expect the others to do the same),

* * * * * * * * * * * * *

4 Once again, length constraints forbid offering a full account of the prisoners' dilemma game. However, this game is popular enough to be found in almost any book related to game theory.

what properties must the 'solution' possess?" The answer, adopted in this paper, is that the Nash equilibrium concept is the necessary, yet not sufficient, condition for an outcome to be a solution.

II. Resolving the Infinite Regress Problem

The previous section concluded by asserting that the solution to a game must be a Nash equilibrium. This section introduces the concept that game theory and, subsequently, economic theory have been employed to prove the existence of equilibria. I then proceed to offer a general style (general approach) for solving infinite regress, while introducing the associated mathematical notions.

Given a noncooperative game G=(I,S,U), where I is the set of players, S the vector of strategy spaces, and $U:\Pi(S) \rightarrow R$ a utility function for each agent, we can define a best reply map for all i's in I.[5] The best reply map provides a natural relation to equilibrium points. An equilibrium point must be a best reply to itself, and any strategy combination that is a best reply to itself must be an equilibrium point. The following lemma formalizes this concept.

1. Lemma: (Friedman [1986] p. 36) Let G=(I,S,B) be a noncooperative game. $s \in S$ is an equilibrium point of G iff $s \in f(s)$, i.e., s is a fixed point of f.

So, given the mathematical specification of a game, the problem of solving for the equilibria reduces to the existence of fixed points.

* * * * * * * * * * * *

5 The best reply mapping for player i is defined as a relationship associating each strategy combination of $s \in S$ with $s_i \in S_i$ according to the following rule: $f_i(s)=\{t_i \in S, U_i(s\backslash t_i)=\max U(s\backslash s'_i)\}$. The best reply mapping is $f(s)=f_1(s) x \ldots x f_n(s)$.

Conditions for uniqueness, stability, and parametric dependence of the fixed point are sought thereafter.[6]

Let us now investigate a bargaining game. Suppose we have two agents (1 and 2), and one perfectly divisible good (X). Agents have identical utility functions, and without loss of generality we let $U_1(x)=U_2(x)=x$. The players must agree on a division of the good. They simultaneously announce a demand $D(x^i) \in R^+$, with i=1, 2. If agreement is achieved, the agreed-upon division is implemented. If no agreement is achieved, the good is evenly split between the players: 1 gets X/2 and 2 gets X/2. For the moment, we assume the players know each other's utility functions. Therefore, players know each other's reaction function $f:R^+ \dashrightarrow R^+$; in addition, the players know that f is common to both. The fixed point solution of this game is rather trivial because

$$D(x^i)=f(D(x^j))$$
$$\text{and } D(x^j)=f(D(x^i)),\ \text{with } x^i=x^j \geq X/2.$$

Let us complicate the situation by relaxing the common knowledge assumption and by postulating that if no agreement is achieved, then neither player receives anything, so $U_1=U_2=0$. Now players have an incentive to coordinate their announcements, by suggesting mechanisms that induce cooperative play. However, equilibrium expectations are too important to ignore. Agents do not have the knowledge of each other's reaction functions, so 1 (2) cannot decide on a mechanism unless he forms beliefs on 2 (1)'s reaction function over

* * * * * * * * * * * *

6 By stability, I mean that starting with an arbitrary initial point, the system will converge to the fixed point.

mechanisms. In essence, 1 (2) must assess the probability that 2 (1) agrees to a given mechanism. Let S be the set of all possible mechanisms. Let P(S) be the set of probability measures on S, such that for $\forall \ s \in S$, s is assigned a probability p(s). So 1's beliefs belong to $D_1(x^1)=P(S)$, and 2's beliefs belong to $D_1(x^2)=p(S)$. Yet, 2 (1)'s reaction function is itself a function of his beliefs. So 1 (2) needs to form beliefs on 2 (1)'s beliefs about 1 (2)'s reaction function. Thus:

$$D_2(x^1)=P(S\times D_1(x^2))$$
$$\text{and } D_2(x^2)=P(S\times D_1(x^1)).$$

Proceeding in this way, we get a system of difference equations:

$$D_{t+1}(x^1)=P(S\times D_t(x^2))$$
$$D_{t+1}(x^2)=P(S\times D_t(x^1)),\ t\geq 1$$
$$\text{with } D_1(x^i))=P(S) \text{ for } i=1,\ 2.$$

Clearly, $D_t(x^1)=D_t(x^2)=D_t(x)$, $t\geq 1$, and we can write

$$D_{t+1}(x)=P(SxD_t(x))=F_t(D(x)).$$

We can interpret each $F_j(D(x))$ as an attempt at coordinating actions in $F_{j-1}(D(x))$, that is, $F_j(D(x))=F(F_{j-1}(D(x)))$. In this formulation, the players' suggestions about mechanisms are a function of their beliefs about each other's reaction mapping. Using arrows, we can represent the system diagramatically:

2. Diagram:

$$F_0(D(x))\longleftarrow\ F_1(D(x))\longleftarrow\ F_2(D(x))\longleftarrow\ \ldots\longleftarrow\ F_n(D(x))\longleftarrow\ F_{n+1}(D(x))\ldots$$

with $F_0(D(x))$ being the original game played noncooperatively.

Unfortunately, traditional analysis cannot provide a solution for this system of difference equations. Consequently, traditional fixed-point theorems cannot be invoked. First, the left-hand side is a set of points in R^+, while the right-hand side is a set of probability measures. Second, the domain of F cannot be restricted to be a set anymore, since for a fixed point of F to exist, F must map sets (rather than the elements between them). So the domain of F must be the collection of all sets, which is not a set, by the Russell paradox. The mathematical tool of categories provides a proper setting.[7] Indeed, as will be seen later, category theory provides appropriate notions of continuity, limits and fixed points.

3. Definition: A *category* K is defined by

i-a class of *objects*: x,y,...; denoted by obj(K).

ii-a class of *arrows (morphisms)* between those objects: f,g,...; denoted by K(x,y) for each x and y in obj(K).

An associative operation called *composition* that associates to each pair of morphisms f:x--->y, g:y--->z a morphism fg:x--->z, and for every object x, a morphism id_x:x--->x, the identity on x, such that fid_x=f and $id_x g$=g.

Note that the object class of a category provides a setting for the domain of F (defined in section III as a collection of strategy sets). We can define a structure preserving relation between categories.

* * * * * * * * * * * *

7 For a rigorous treatment of category theory, see Arbib and Manes (1975) and Mac Lane (1971).

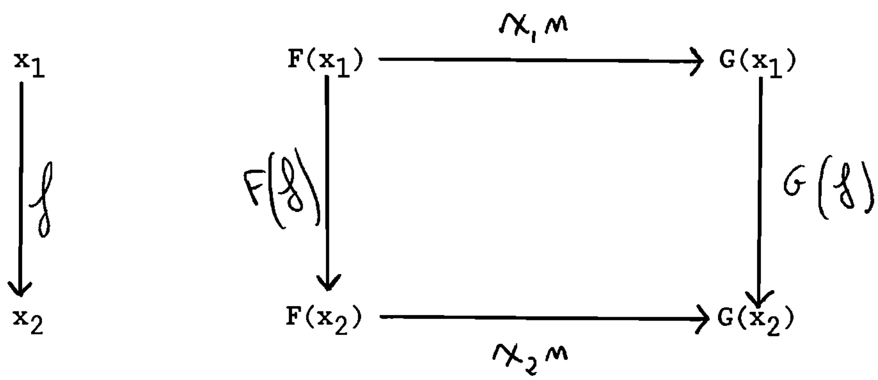
4. <u>Definition</u>: Given two categories K and C, a *functor* F:K--->C assigns to $\forall x \in obj(K)$ $F(x) \in obj(C)$, and to each morphism $f \in K(x_1,x_2)$ a morphism $F(f) \in C(F(x_1),F(x_2))$, in such a way that composition and identity are preserved:

$$F(fg)= (Ff)(Fg)$$

$$F(id_x)=Fid_x.$$

We can think of a functor F as giving a representation of K in C. Now we introduce a concept that translates the representation F to another representation G:K--->C.

5. <u>Definition</u>: Given two functors F:K-->C and G:K-->C, a collection of morphisms in C <xn: xF-->xG with $x \in obj(K)$> is a *natural transformation* from functor F to G if for all x_1 x_2 in obj(K) and for any $f \in K(x_1,x_2)$ the following diagram commutes, that is, if different paths yield the same overall function.



I concentrate on (right and left) chains in a category C, then study their relation to fixed-point concepts. A *right chain* is an arbitrary sequence $<c_n/n\geq0>$ of morphisms of the form

$$C_0 \longrightarrow C_1 \longrightarrow C_2 \longrightarrow \ldots.$$

6. <u>Definition</u>: Given a category K we define its *dual (opposite)* category K^{op} by

$$obj(k^{op})=obj(k)$$

$$K^{op}(a,b)=\{b\dashrightarrow a: f \in K(a,b)\}$$

with composition defined by c-->b-->a=c-->a; identities are the same as K with arrows reversed.

A *left chain* in C^{op} is a right chain in C.

7. <u>Examples</u>:

We can readily think of a category whose morphisms are left chains. Let W be the category with

i-obj(w)=N (all natural numbers).

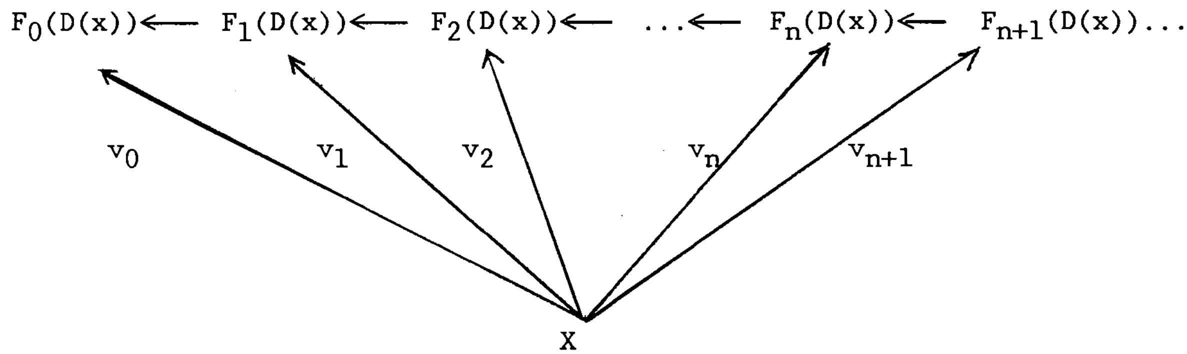
ii-$\forall j\&i \in obj(w)$, if $j \leq i$, $\exists$ exactly one arrow from i to j (there are no other morphisms).

If i-->j-->k, for i,j,k $\in$ obj(w), define composition to be the unique arrow i-->k. Identity arrow is i-->i, for all $i \geq 0$.

The dual W^{op} is obtained simply by changing the order of arrows: if $j \geq i$, there is exactly one morphism from j to i, $i \geq 0$.

Diagram 2, developed earlier, is an example of left chains. If we can construct an X such that the above diagram commutes at every level, then we would have a candidate for a universal mechanism. In diagrammatic form, we wish to construct

8. <u>Diagram</u>:

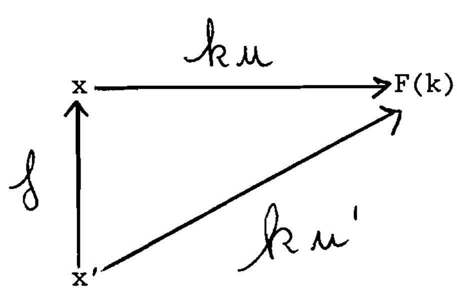


with all triangles commuting.

If $v_0 \ldots v_n$ are natural transformations, then X would be the limit of functor F. Under the proper specification, X turns out to be the desired fixed point. We have motivated the following definitions.

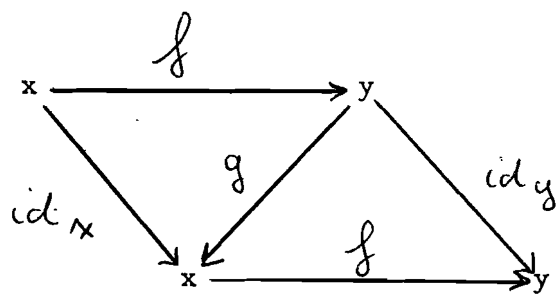
9. <u>Definition</u>: A *constant functor* I_c:K--->C is a functor that assigns to $k \in obj(K)$ the same $c \in obj(C)$, and to each morphism in K the identity morphism id_k on k.

10. <u>Definition</u>: Let F:K--->C be a functor. A limit of F is an object $c \in C$ and a natural transformation u: I_c--->F with the following universal property: If $c' \in obj(C)$ such that $c' \neq c$ and u':I_c--->F is any other natural transformation, $\exists$ a unique morphism f:c'--->c in C that makes the following diagram commute $\forall$ $k \in obj(K)$.



Recall that the objective in this section has been to show how the concepts of limits and fixed points can be generated outside the realm of traditional analysis, which can be built only on well-founded sets. So far, a definition of these terms has been offered using categories (objects and arrows) and functors. The relation between the limit and the fixed point(s) of a functor is formalized by the generalized Kleene fixed-point theorem. First, we must introduce some new concepts.

11. Definition: Given $x \& y \in obj(K)$, $f:x \dashrightarrow y$ is said to be an *isomorphism* if $\exists$ a $g:y \dashrightarrow x$ such that $fg = id_y$ and $gf = id_x$. In diagrammatic form,



commutes.

12. Definition: Given a functor $F:K \dashrightarrow K$, an object x of K is a *fixed point* of K if $\exists$ an isomorphism $f:x \dashrightarrow F(x)$ in K.

13. Definition: A *terminal object* in a category K is an object denoted by 1 such that $\forall\ x \in obj(K)$, $\exists$ a unique morphism $!:x \dashrightarrow 1$. An *initial object* in K is a terminal object in K^{op}.

14. Definition: A functor $F:C \dashrightarrow K$ is *continuous* if whenever $c_n:C_n \leftarrow C_{n+1}$ is a left chain in C and (U,u) is a limit for $\langle c_n \rangle$, then (FU,Fu) is a limit for $\langle FC_n \rangle$.

15. Generalized Kleene Fixed-Point Theorem: Let K be a category with an initial object 1 such that every right chain has a colimit. Then every continuous functor F:C--->C has a least fixed point the colimit of the right chain

$$\ldots \longrightarrow F^2(1) \longrightarrow F(1) \longrightarrow 1.$$

Proof. (see Manes and Arbib [1986], p. 270).

Note that the Kleene fixed-point theorem applies to right chains. But as indicated in diagram 8, we are interested in the limit of a left chain. Theorem 16 establishes a fundamental result of category theory that allows us to state the dual of the fixed-point theorem without a reference to any proof.

16. Duality Principle for Category Theory (Arbib and Manes [1975]):

Let T be any construct defined for any category K. Then the dual of T, called coT, is the construct defined for any category K by defining T in K^{op} and reversing all arrows.

If T is a theorem true for all categories K, then the dual of T, obtained by reversing all the arrows of T, is true for all categories K^{op}, and thus (since $(K^{op})^{op}=K$) is true for all categories.

17. Dual: Let K be a category with a terminal object D(x) and such that every left chain has a limit. Then every continuous functor F:K--->K has a greatest fixed point the limit of the left chain

$$F_0(D(x)) \longleftarrow F_1(D(x)) \longleftarrow F_2(D(x)) \longleftarrow \ldots \longleftarrow F_n(D(x)) \longleftarrow F_{n+1}(D(x)) \ldots .$$

Stability of the fixed point is the result of it being an approximation from the iteration process represented by the left chain. The uniqueness of the greatest fixed point is settled by the following results.

18. Theorem: Limits are unique up to isomorphism.

19. Corollary: The greatest fixed point of F:K-->K is unique up to isomorphism.

Note that F may have other fixed points; however, they cannot be obtained as a limit of the same chain.

20. Proposition: This specification guarantees that the fixed point (if it exists) is continually dependent on the parameters of the game.

Proof. As a limit X is contingent on the underlying game by the construction of the left chain.

Manes and Arbib (1986) present a reference for the adaptation of categorical techniques to computations of data types. Vassilakis (1989, 1990) provides categorical constructions relevant to economic theory.

III. A "Revised" Nash Demand Game

The previous section showed how the existence of a universal mechanism reduces to the existence of the fixed point of a functor. I have not explicitly constructed that functor; nor have I determined its proper domain and range categories. This section presents an adaptation of the universal mechanism constructed in Vassilakis (1989). The underlying game consists of the Nash demand game. Agents simultaneously announce a demand vector $x_i \in X_i$, the dimension of which is determined by the number of traded goods in that economy. Every player has an initial endowment e_i. A von Neuman-Morgenstern utility function is defined. If the demand matrix $x=(x_1 \ldots x_n)$ is such that $\Sigma x_i \leq \Sigma e_i$, i.e., markets clear, then every agent receives his demand, thus a utility of $U_i(x_i)$. Otherwise, if $\Sigma x_i > \Sigma e_i$ then the game is played noncooperatively: every player consumes his endowment and receives a utility of $U_i(e_i) \leq U_i(x_i)$. The *Nash equilibria* form the set of demand matrices yielding Pareto-efficient outcomes.

This specification of Nash equilibria eliminates suboptimal equilibria similar to the equilibrium arising in the prisoners' dilemma game. However, the game has multiple equilibria. Every matrix x with the specification $\Sigma x_i = \Sigma e_i$ is Pareto optimal and hence an equilibrium. Players can suggest different mechanisms to chose among the equilibria by playing a game over mechanisms. Yet, as argued earlier, it must also be shown that the latter game yields a solution. Eventually, we are led to an infinite regress. The construct, presented in Vassilakis (1989), considers an aggregate revision functor that gives every player the power to suggest a new outcome at every level of play. Recall that the objective is to extract the existence of a fixed point of a functor over mechanisms.

Let I be the following category:

Obj(I)=players, denoted by an integer $i=1,\ldots,n$

$K(i,i)=id_i$, for all i in obj(I). There are no other morphisms.

Categories with only identity morphisms are called discrete.

Let a collection of strategy spaces be a functor S:I-->K, where K is a category whose objects are abstract sets and whose morphisms are abstract input/output programs. A given S(i) specifies player i's strategy space. One of the benefits of choosing category I to be discrete is that it enables us to define the aggregate strategy space as $S_1 \times S_2 \times \ldots S_n = \Pi\ S_i$ (× denotes the product).

I define a mechanism to be a triple (S,f,O), where $S = \Pi\ S_i$, $O \in obj(K)$ is an outcome space, and $f: \Pi\ S_i \dashrightarrow O$ is a morphism in K. A primitive mechanism can be depicted as a triple (A,a,R^n) with $A \in obj(K)$ such that

$A = \Pi\ A_i = \Pi\ [0, \Sigma e_i\ ]$;

$a: \Pi\ [0, \Sigma e_i\ ] \dashrightarrow R^n$,

$$a(x_1,x_2,\ldots,x_n)=(x_1,x_2,\ldots,x_n) \text{ if } \Sigma x_i \leq \Sigma e_i$$
$$\text{and } (e_1,e_2,\ldots,e_n) \text{ if } \Sigma x_i > \Sigma e_i.$$

Here the primitive mechanism denotes the play of the Nash demand game (the underlying game) without cooperation, capturing all the multiple equilibria. In order to relate the ideas developed so far, the next section introduces some new concepts.

<u>Definition</u>: The *category of functors from I to K*, denoted by [I-->K], has as objects all functors F:I--->K and as morphisms all natural transformations between them. In this instance I call --> the *functor space constructor*.

<u>Definition</u>: Given categories K and C, I construct the *product category* K×C as follows: $<k,c> \in obj<K \times C>$ if $k \in obj(K)$ and $c \in obj\ (C)$. A morphism of K×C is a pair <f,g,> with f a morphism in K and g a morphism in C. Composition is defined in terms of the composites in K and C:

$$<f',g'><f,g,> = <f'f,g'g>.$$

<u>Definition</u>: A *coproduct* is a colimit of a functor F:I--->K on a discrete category I. For example, in the category **Set**, whose objects are all sets and whose morphisms are the inclusion map, coproduct is the disjoint union. The coproduct of two objects is denoted by "+".

<u>Definition</u>: A *polynomial functor* F:K--->K is a functor that can be constructed from constant or identity functors through the use of products, coproducts and compositions.

Each agent must be endowed with the capability of either proposing to coordinate in a given mechanism or proposing to coordinate on the proposals in that mechanism. So an agent's revision functor is defined as

R_i:[I-->K]×K--->K×K by $R_i(S,O)=(S',O')$, where

$$O' = [\Pi\ S_i \text{--->} O] \text{ is a new outcome space}$$

$$\text{and } S' = O' \times [O' \times \ldots \times O' \text{-->} O']$$

is a new strategy space for each player i. The revision functor defines what each agent can propose. This definition captures the fact that given a game with strategy space S and outcome space O, each agent i can simultaneously make two proposals. The first proposal is a mechanism $f \in O'$, that is, i's proposal on how to coordinate actions on the original game (S,O). The second proposal is a mechanism that selects one out of n proposals in O'.

The aggregate revision functor

$$R:[I\text{-->}K]\times K\text{--->}[I\text{-->}K]\times K$$

is defined by $R(S,O)=(A,R^n)+(\underline{S}',O')$, where $\underline{S}':I\text{-->}K$ with $\Pi\ \underline{S}'(i)=S'$; S' and O' are defined as above. A universal mechanism (S,f,O) must be a fixed point of that functor: $(S,O)\sim(A,R^n)+R(S,O)$. The meaning of the fixed point equation is that a strategy is either primitive or a revision strategy for each i in obj(I), where $S_i\sim A_i+[O'\times\ldots\times O'\text{-->}O']$ with $O'=[\Pi\ S\text{-->}O]$, and that an outcome is either primitive or an outcome of the revision mechanism $X\sim R^n+[\Pi\ S_i\text{--->}O]$.

To satisfy the statement of the functorial fixed-point theorems, a a category K is needed that satisfies the following:

1. K has limits of the left chains.
2. K has a terminal object.
3. K has polynomial functors, all of which are continuous.
4. K has a functor (called a function space constructor) -->: K×K--->L defined on two objects x and y in obj(K) by -->(x,y)=[x-->y], such that --> is continuous. Note that the morphism space does not belong to K but to a larger category L of which K is a subcategory. In other words, L is the category with the desired function space. However, L has too many morphisms

for [L-->L] to be made into a functor. Thus, I restrict the set of morphisms in L to K and obtain functoriality (see Manes and Arbib [1986], p. 313).

To state the main theorem of this section, again, some new concepts must be identified. I turn to partially ordered sets (posets). A poset can be seen as a specialization of a category that allows for much insight about general categories with minimal loss of generality. When posets are regarded as categories, a monotone map (function) represents the same concept as a functor.

Definition: A partially ordered set, *poset*, is a pair (P,R) where P is a set and R is a binary relation on P, which is a partial order on P. Then the following axioms hold:

i- Reflexivity: xRx

ii- Transitivity: $xRy \wedge yRz \rightarrow xRz$

iii- Antisymmetry: $xRy \wedge yRx \rightarrow xEy$.

Note that posets are themselves categories; examples are W and W^{op}.

Definition: A poset is a *domain* if it has a least element and if whenever $(x_n: n=1,2,3,\ldots)$ is an ascending chain in P (i.e., $x_n \leq x_{n+1}$), then a least upper bound (LUB) $\{x_n\}$ exists.

Definition: Given domains D and D', let [D-->D'] be the set of all continuous functions f:D-->D' partially ordered by $fRg \Leftrightarrow f(x)Rg(x),\ \forall\ x \in D$.

Proposition (Manes and Arbib [1986]): [D-->D'] is a domain under R. I call [D-->D'] a *function space domain*.

Definition: Given domains $D_1,\ldots,D_n$, define i- $D_1 \times \ldots \times D_n = (x_1, \ldots, x_n)$ with $x_i \in D$, $(x_1, \ldots, x_n) \leq (y_1, \ldots, y_n)$ iff $x_i R y_i \in D_i$.

ii- $D_1 + \ldots + D_n = \{\perp\} \cup \{1\} \times D_1 \cup \ldots \cup \{n\} \times D_n$, where $\perp \leq z$, $\forall$ z in $D_1 + \ldots + D_n$, while (i,x)R(j,y) if i=j and xRy in D_i.

<u>Proposition</u>: $D1 \times \ldots \times D_n$ and $D_1 + \ldots + D_n$ are both domains.

<u>Definition</u>: *Categories of Domains*

i- Let Dom_c be the category with

obj(Dom_c)=domains

and K(D,D')=continuous maps.

ii- Let Dom be the category with

obj(Dom)=domains

and K(D,D')=strict maps.

iii- Let Dom_{adj} be the category with

obj(Dom_{adj})=domains

and K(D,D')=maps having an adjoint.

<u>Remark</u>: Dom_{adj} is a subcategory of Dom_c.

<u>Fact</u>: Both Dom_c and Dom have limits (colimits) of left (right) chains, as well as an initial (terminal) object.

<u>Fact</u>: Any polynomial functor Dom_c-->Dom_c is co-continuous and therefore has a least fixed point.

Proof (sketch). Constant functors and the identity functor are co-continuous, and so is any composition of co-continuous functors. The product of co-continuous functors is continuous in closed categories. The coproduct of co-continuous functors is also co-continuous.

The necessary and sufficient resources are now in place to provide a category that can capture the universal mechanism I have, so far, been seeking.

<u>Theorems</u>:

i- D--->[D-->D] extends to a functor Dom_{adj}-->Dom_{adj}.

ii- The terminal object in Dom is terminal Dom_{adj}.

iii- Given a left chain in Dom_{adj} with limit (a,n) in Dom, it follows that (a,n) is its limit in Dom_{adj}.

iv- The functor [D-->D] is continuous.

v- Every polynomial functor Dom-->Dom maps Dom_{adj} into Dom_{adj}.

Proof. Manes and Arbib (1986), pp. 311-317.

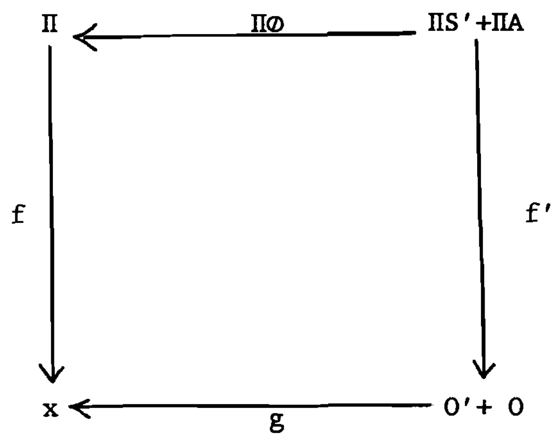
Indeed, by inspection Dom_{adj} satisfies *desiderata* 1-4 enumerated earlier in this section. However, I have not yet shown the existence of the fixed point.

<u>Corollary</u>: (Vassilakis [1989]) A universal mechanism exists.

Proof. (S,O) is defined as a fixed point of a functor defined by the continuity-preserving operations on left-chain functors (the product operation preserves continuity). The outcome function f: Π S_i--->X is arbitrary.

It must still be shown that the transformation of the proposals into outcomes is well defined.

<u>Corollary</u>: (Vassilakis 1989) If (S,f,O) is a universal mechanism, then there exists a unique outcome function f': Π S'_i+A-->O'+R^n (with S' and O' defined as above) that is consistent with f, in the sense of making the following diagram commute:



where Φ:S'+A-->S and g:O'+R^n-->O are the fixed-point isomorphisms.

Proof. f'=g^{-1}(f(Φ)) is well defined. Note that f' transforms proposals into outcomes.

Hence, f' extends uniquely to an outcome function on the proposed revisions of f. This completes the specification of the "revised" Nash demand game.

So far, I have defined f to be an input/output morphism of K (a function of Dom_{adj}). An explicit specification of f is tantamount to an explicit specification of the game, but this must be addressed in future research.

IV. Conclusion

This paper has attempted to show how category theory provides the tools for solving the infinite regress that appears in the statement of many social-choice theories. A revision functor was constructed and specified on the Nash demand game. Solving the "revised" Nash demand game is beyond the scope of the present project and will have to be explored in another paper. The aim here was simply to outline the main difficulties facing social-choice theories and to show how categorical tools can be used for constructing the appropriate models by providing a setting for the concepts of continuity, limits, and fixed points.

The discontent with moral philosophy has been articulated by Williams (1985). In his account, the difficulties are rooted in the fact that modern morality theories are "...governed by a dream of a community of reason that is too far removed..." (p. 197). In our case, the exogenous bargaining procedures and the "nice" properties they must have in order for the Gauthier results to go through are essentially the embodiment of that community of reason. The novelty in this paper stems from its ability to provide a construction that escapes the analyst's biases.

Other pertinent problems can be settled by the same techniques; for example, bounded rationality and universal beliefs spaces (see Mertens and Zamir [1985]). While there remains much to be done to expand the boundaries of moral philosophy, I hope to have offered a promising approach.

Appendix

This game illustrates how information leakages restrict the set of feasible mechanisms. In this game, agents' information sets are not identical. I define an *incentive-efficient* mechanism as an incentive-compatible mechanism such that there is no other incentive-compatible mechanism that is at least as good for all agents, and strictly better for at least one agent, in the truthful equilibrium. The term *ex ante* refers to a situation in which agents have not yet observed their types; *interim* refers to the case in which agents have observed only their types and not other agents' types.

Consider the two-agent economy, where an agent is either type a or type b with equal probability. Suppose we have three decisions {A,B,C}. The utility of each agent under every decision is self regarding and a function of his respective type as represented in table 1.

Table 1

	u_{1a}	u_{1b}	u_{2a}	u_{2b}
d=A	2	0	2	2
d=B	1	4	1	1
d=C	0	9	0	-8

Let D be the following decision rule:

$$D(1a,2a)=A, \quad D(1a,2b)=B, \quad D(1b,2a)=C, \quad D(1b,2b)=B.$$

Decision rule D selects C only if agent two's type is la. Otherwise, if two is type 2b, then only decision B is chosen. Rule D is incentive compatible, which can be confirmed by inspection. For types 1a, 1b, and 2b, it is trivial to show that an honest reporting is the dominant strategy. Type 2a can either get B with probability one, by actually reporting his false type, or get A or C with equal probability by honestly reporting his type. Since the expected utility of both prospects is identical for type 2a, he is willing to report his type honestly when D is implemented (under risk neutrality). Decision rule D is interim incentive-efficient. However, if agent one knows that he is type 1a, then he knows that both he and player two prefer decision A over the outcome proposed by decision rule D. Thus, agent two would expect (with probability one) agent one to call for decision A if one was type 1a. If agent one insists on D then agent two can infer that type one is 1b. In this event, agent two is better off reporting 2b, regardless of his true type, in order to avoid decision C altogether; decision D's incentive-compatibility property is thereby destroyed. Because of its simplicity, this example does not show an even stronger case where an ex-ante incentive efficiency is not interim incentive-efficient; however, it should be clear that such a case is possible.

References

Arbib, M., and Manes, E., Arrows, Structure and Functors: The Categorical Imperative, Academic Press, New York, 1975.

Crawford, V., "Efficient and Durable Decision Rules: A Reformation," Econometrica, Vol. 53, No. 4, July 1985.

Friedman, J., Game Theory with Application to Economics, Oxford University Press, New York, 1986.

Gauthier, D., "Justice as Social Choice," in Morality, Reason and Truth, D. Copp and D. Zimmerman, eds., Rowman and Allanheld, Totowa, New Jersey, 1984, pp. 251-269.

Holmstrom, B., and Myerson, R., "Efficient and Durable Decision Rules with Incomplete Information," Econometrica, Vol. 51, No. 6, November 1983.

Kreps, D., A Course in Microeconomic Theory, Princeton University Press, Princeton, New Jersey, 1990.

Mac Lane, S., Categories for the Working Mathematician, Springer-Verlag, New York, 1971.

Manes, E., and Arbib, M., Algebraic Approaches to Program Semantics, Springer-Verlag, New York, 1986.

Mertens, J.F., and Zamir, S., "Formulation of Bayesian Analysis for Games of Incomplete Information," International Journal of Game Theory, Vol. 14, Issue 1, 1985.

Rawls, J., A Theory of Justice, The Belnap Press of Harvard University, Cambridge, Massachusetts, 1971.

Vassilakis, S., "Economic Data Types," Working Paper No. 252, University of Pittsburgh, 1989.

____________. "Functorial Fixed Points. A Non-Technical Introduction," Mimeo, University of Pittsburgh, 1990.

Williams, B., Ethics and the Limits of Philosophy, Harvard University Press, Cambridge, Massachusetts, 1985.