# GAME THEORY AND THE INTERPRETATION OF DEONTIC LOGIC

LEO APOSTEL

Several uninterpreted formal systems have been proposed as expressing the syntactical properties of concepts like 'obligation' or 'permission'. No agreement has been reached and divergences have been important [1]. This is not astonishing, because it is very difficult to discuss the adequacy of a formal system, if the intuitive modal, whose properties the formalism wants to mirror, is not clearly or not uniquely defined.

Time has come, however, to look for such a model. Far from us the idea to criticize the attempts made to start the formalisation without clear interpretation. Formalism and model reciprocally clarify each other; without the systems that have been syntactically described, the search for a model would be hopeless. But now a sufficient number of attempts lay before us. We can proceed to the application of the maxim: only through its interpretation is an algebra a logic; and only through its interpretation is a modal algebra a deontic logic.

We shall try to perform essentially two different tasks: 1. We shall start with von Wright's deontic logic, interpret his variables for acts, his combinators of acts and his modal operators by means of game theoretical concepts. Our aim here will not be to argue that this is the only or best interpretation for deontic logic; we only wish to stress that such an undertaking is possible and that for different, equally plausible game theoretical interpretations the formal properties of the deontic operators will show variations (this will perhaps

[1] See H. L. von Wright, *Essay on Modal Logic*, Amsterdam, 1951,

A. Prior, *Formal Logic*, Oxford, 1955; *Time and Modality*, Oxford, 1956.

R. Feys, *Expression modale du «devoir être»*, Journal of Symbolic Logic, 1954, p. 91-92.

H. N. Castañeda, *La Logica general de las Normas y la Etica* (Universidad de San Carlos, Guatemala, n° 30, pp. 129-196); *Un Sistema general de logica normativa* (Dianoia, Mexico City, vol. 3, 1957, pp. 303-333); *The Logic of Obligation. Philosophical Studies*, February 1959, pp. 17-23.

O. K. Moore and A. Ross Anderson: *The Formal Analysis of Normative Systems*, Technical Report n° 2, Contract n° SAR/609(16), Office of Naval Research/Group Psychology, 1956.

prepare an understanding of the disagreements that have been current).

2. In a second section of this paper, we shall discuss the adequacy of the different game theoretic concepts as models for ethical norms. Here our aim will be to point out some models as more adequate than others; naturally the most adequate models will be the most complex ones. We shall be unable yet to study them syntactically; but our remarks will show the possibility of such a study.

## Section I

1. Let us describe a game G [2].

A game G includes:

1) a set N of players,

2) a set X of positions,

3) For every member $x_j$ of X and for every member $n_i$ of N, an order of preference $R_j$ $(x_j)$, indicating where the position in question is situated in $n_i$'s preference ranking,

4) An application A from X to X, showing what positions can follow what other positions,

5) An application from X to N, stating what players determine what changes of position,

6) The set P of all plays (meaning; the set of all sequences starting with an x that has no A-predecessor, finishing with an x that has no A-successor, and containing only members obtained from each other by operations admitted by A).

7) The set S of all strategies: a strategy is application of the set of all positions, controlled by one player, to itself, selecting among the A successors of each of these positions, one and only one member for every play determined by its initial position $x_0$.

8) The information function I is an application of the elements of (X, N) (set of sets of positions associated with one player) to sets of elements of (X, N) determining for every position controlled

[2] For elegant and clear introduction to mathematical game theory, see J. C. Mac Kinsey *Introduction to The Theory of Games*. An excellent review of the content of game theory, essential for us, through the wealth of references to ethical concepts is Duncan Luce and Howard Raiffa *Games and Decisions* (1957, Wiley). A recent, extremely algebraic approach, useful for establishing relationships with logic is Claude Berge *Théorie Générale des Jeux à n Personnes,* Mémorial des Sciences Mathématiques, fascicule CXXXVIII, 1957.

by a player, the set of all positions that are not distinguished from this given one by this player.

We have tried, following Berge's example, to give definitions as general as possible, because, after all, our intention is now to project this structure on the theory of modal operators.

2. The theory of deontic modalities sometimes (von Wright) refers to acts, sometimes (Feys-Prior) refers to propositions about acts. In both cases we need the concept of 'act'. Let us then begin by looking for a model of 'act' in the formal structure G.

An act associates with an initial situation and an agent, a final situation provoked by that agent. Situations are here the members $x_i$ of the set X of positions, agents are here the players. An act then would be here a move in a play, i.e.: for every member $p_x$ of P, and every member $n_i$ of N, an act is an application of (X, p, n) on (X, P, N). If then a strategy is a general rule of action, an act is the application of a strategy to a particular position.

However, on p. 36 of his 'Essays in Modal Logic', von Wright stresses that he wants to use this term 'act' not for an individual act, but for a property ('theft' is called an act). In our game theoretic model, this would equalize 'act' with a class of moves, or even with a general strategy itself (a strategy is certainly a class of moves). It seems however that we would be untrue to von Wright's intentions, if we wanted to restrict 'acts' to 'strategies'; we should leave open the way in which the class of moves, corresponding to acts, is defined. Theft is an action, that does not refer to a particular initial situation, nor to a particular final situation, but only to the relation between both. But such a concept does not refer to a complete sequence of moves stating in every situation what to do. A strategy would correspond to a way of life, an act to a class of moves (parts of the same or of different strategies) or even to a class of strategies (namely the class of strategies containing certain moves or exhibiting specific relations between those moves).

Propositions about acts state that certain acts have been performed; we can thus characterize them as members of information sets, communicating that certain players (or sets of players) perform certain moves (or sets of moves).

Let us hasten to remark that we can conceive of games on games, and that here the choice of a preference function, or its modification, the choice of a strategy or its modification, the selection itself of players and positions of the game can all be moves or classes of moves in higher order games. In fact it may appear that most acts mentioned in moral rules belong to this higher level. But it will be evident that a theory about these higher order acts can only be developed after knowledge has been obtained about lower order acts.

Let us then, for our present purposes, adopt the following definition 'acts are classes of applications of strategies to positions'.

3. Having defined acts as classes we are certainly allowed to consider Boolean functions of acts (because we are allowed to consider union, intersection, complement and inclusion of classes). However acts are, according to the definition adopted, very peculiar classes; they are classes of $( s_i p_j x_k n_l )$ complexes. It seems then advisable not only to state automatically that Boolean functions can be defined, but to examine the multiplicity of methods that can be followed to define Boolean relations on games and on constituents of games.

Let us examine this multiplicity with reference to the inclusion relation.

All other properties being constant, we might say that a game is included in another game $( G1 \subset G2 )$ in one of the following circumstances:

*a.* The set of players of the one is a subset of the set of players of the other: $N1 \subset N2$.

*b.* The set of positions of the one is a subset of the set of positions of the other: $X1 \subset X2$.

*c.* The preference ranking of the one is a refinement of the preference ranking of the other (meaning that whenever there is a preference in R1, it exists in R2 but there are cases of equivalence in R1 that are no longer equivalences in R2).

*d.* The codomain of the operator A1 is a subset of the codomain of the operator A2 : all positions that are allowed to follow a given position in A1 are also allowed to follow that position in A2, but there are allowed positions following the given one in A2 that have not this characteristic in A1.

*e.* The information sets of the first game are super sets of the information sets of the second game; all acts that are confounded with each other in the second game are also confounded in the first game, but there are also alternatives undistinguishable in the first that are distinguished in the second.

*f.* Several of the five preceding alternatives can be combined with each other: a large number of new meanings must thus be added to the ones we already have considered (we think that the following two cases: the class of strategies of the first game is a subset of the class of strategies of the second game, and the class of plays of the first game is a subclass of the class of plays of the second game, are such complex cases).

In one word: the inclusion of games reduces to the inclusion of one or more of the constituents of those games. This is not only the case for inclusion, but obviously it is also the case for union, intersection and complement. The fact that this has already been clearly recognized by more than one writer is attested by the following two facts: 1. Gale and Stewart have defined the concept of subgame of a game in a way reminding us of *d*, and 2. Berge has defined the complement of a preference ranking, by choosing a special case of a definition akin to *(c)*.

The study of general laws about these Boolean functions of games would certainly be most necessary and rewarding. We cannot do more here than attract attention to it, because our aim is now to come to the definition of Boolean functions of acts.

An act, as the application of a strategy to a position (or rather as a class of such applications) refers to a subgame of a game. All Boolean functions for games can ipso facto become Boolean functions for acts.

An act a can thus be included in another one b: (a) because the class of positions to which the common strategy is applied is a subclass of that other one, (b) because the resulting position of the first is part of the resulting position of he second, (c) because the set of actors of the first act is a subset of the set of actors of the second set d. because the first game confounds more positions than the second one, within the same group of positions. Some combination of these possibilities can equally be considered.

It is easy to recognise the complete analogy between the definition of these Boolean functions for games, and the corresponding Boolean functions for acts.

An act will be said to be in the *complement* (i = 1, 2) of another act, if either the application of both together is impossible (c1) or the one is the inverse of the other (if applied to the final position of the first, the final position of the second is the initial position of the first) (c2).

These two, already different features, can be realised in a great variety of ways, through conditions on the positions, the players, the preference rankings and so forth. Once more, the same multiplicity as the one encountered for games as wholes is present here when we want to define Boolean functions for isolated acts.

We do not need to state explicitly the same facts for conjunction and disjunction. Our conclusion is clear: just as in the case of Boolean functions on games, Boolean functions on the applications of strategies (id est: on acts) are definable, but not uniquely definable. We wanted to go into some detail here because in von Wrights approach, Boolean functions on acts are completely reduced to Boolean functions on classes; we do not deny this general remark but want to add that the multiplicity of ways of realization for intersection, union and so forth, in the case of these very special classes that are classes of applications of strategies, should never be neglected.

4. Having defined what is a game, what is an act, and what are the performance functions of acts, we shall now be prepared to define 'obligation' and 'permission'.

Let us first begin with another concept of modal character: an act will be called '*possible*' if it is the class of applications of one of the admitted strategies to one of the admitted positions.

Intuitively surely we want to say that an act is permissible if it can be considered as the application of a strategy such that there is no better one (there may be many equally good). We want to say moreover that an act is obligatory, if it is the only act such that there is no other act equally good or better.

We reiterate here our warning: this is *not* meant as a really adequate interpretation of our terms, it is meant to prepare one. Only for certain specific games, and for certain specific definitions of 'good' will the definitions given here furnish adequate interpretations.

Let us add a second warning: we shall first apply our definitions to strategies as wholes, and later only apply them to acts. This is quite normal, even in classical ethics. An act must be seen as a symptom of a way of life in order to be obligatory or permissible. But we must not neglect the difficulty thus created: the passage from the obligatoriness of a strategy to the obligatoriness of an act is only immediate if the act appears exclusively in a sequence that is an obligatory strategy.

These two warnings being formulated, let us now for the most general case, a non zero sum. n persons game, define what we consider to be a good or best strategy.

We certainly need to define the concept of a good strategy in this most general case where one player does not gain necessarily when the other loses, or loses where the other gains; and where coalitions are possible because more than two players are present.

In traditional game theory, for non zero sum, n person games, an equilibrium point corresponds to the idea of an optimal strategy. An n-tuple of strategies $(s_1......s_n)$ is an equilibrium point if, whenever none of the remaining players changes his strategy, none of the other strategies a given player could choose would yield a higher payoff (more generally: will his result higher up in the preference ranking of final positions). Obviously this definition is a very special one because the possibility of considering joint strategies is not present, and the possibility of pre-play communication to establish mutual compensation is not considered. But let us, for the time being, work with this special definition.

We give then the following definition:

A strategy n-tuple is *permitted* if, and only if, for every $r_j$ and for every $S_i$, M $(s_1......s_n)$ is greater than or equal to M $(s_1...r_i...s_n)$, where M is the evaluation of the outcome. A strategy $s_i$ is permitted with respect to a n-1 tuple if and only if the n-tuple $(s_1... s_i... s_n)$ is permitted.

A strategy is *weakly permitted* if there is at least one n-1 tuple with respect to which the strategy in question is permitted. A strategy is *strongly* permitted if for every n-1 tuple this is the case.

An act is permitted with respect to a given strategy if this act is the application of this strategy to a given position.

An act is *weakly m-permitted* (where m can mean strongly or weakly) if there is at least one strategy that is m-permitted (m having the same values as before) such that this act is permitted with respect to it. An act is *strongly* m-permitted if all strategies that are m-permitted permit this act.

To our mind, the multiplication of concepts that we encounter here is in itself rather enligthening.

Let us now stress that a multiplicity of equilibrium points is in existence for most games.

We could say that an n-tuple of strategies is obligatory if it is either the only equilibrium point, or the equilibrium point with highest payoff (none other with equal or higher payoff being in existence). But the second clause presupposes the possibility of ordering n-tuples of utilities, and interpersonal utility comparison. It seems, as first rough approximation, advisable to use only the first clause.

An act is *strongly forbidden* if for no strategy n-tuple that is an equilibrium point, there is a strategy in it containing this act. An act is *weakly* forbidden if there is in at least one equilibrium n-tuple no strategy including it. An act is *weakly weakly* forbidden if there is in at least one equilibrium n-tuple at least one strategy not including it.

These distinctions being made for «forbidden», similar distinctions should be made for obligation (the maximal equilibrium point, let us not forget it, is a n-tuple of strategies and the subjects of the predicates «obligatory» or «forbidden» are classes of sections of these strategies: the necessity of this passage necessarily introduces the mentioned distinctions).

5. We now study the relations between the two deontic operators and the Boolean functions of acts.

*a.* If an act (or a strategy, or an n-tuple of strategies) is obligatory (in the strong or weak sense), then this act (or strategy or n-tuple of strategies) is permitted. If an act is member of some or all strategies in the unique highest equilibrium sequence, then surely it is member of such strategies in at least one equilibrium sequence. The reverse is not true, as should be the case: there can be equilibrium points without there being a unique or a highest one (highest in the preference order).

*b.* The absence of relation between truth-falsity, and obligation or permission is also realised: there are applied strategies that are not optimal and optimal ones that are not applied.

*c.* The iteration of deontic modalities can have a meaning in games the outcomes of which are other games: this can even happen in more than one way.

Let us sketch two of these possibilities.

[i.] von Wright's «Interpretations of Modal Logic» considers a field of causes, accompanied by effects that in turn can become causes. A state is possible if the presently working causes will produce it; a state is possibly possible if the effects of these present causes will produce it.

Following this lead we could say: a strategy is permissible if it is permissible in GI. It is permissibly permissible if it is permissible in a game that is the outcome of one of the permissible strategies of GI.

A strategy is *obligatorily obligatory* if it is obligatory in each game that is the outcome of an obligatory strategy for GI. In function of this definition, if a strategy is obligatory in a permissible outcome, this does not imply that it is permissible in an obligatory outcome (that might not exist). But obviously OPA implies PPA and moreover POA implies PPA. To develop more theorems one should how-

[3] *Mind,* n. s., vol. 61, 1952, pp. 165-177. Ross Anderson however in his review (JSL vol 18, n 2, p. 177) stresses very well the non topical character of von Wright's interpretations; they have formal similarities but there is not affinity as to content.

ever scrutinize more deeply the relation between the value to a player of a game GI and the values of the games that might be outcomes of GI ([4]).

[ii.] Duncan Luce's concept «psi stability» could here be helpful. For every coalition, only a restricted number of not too different coalitions has, according to Luce, to be taken into account as alternative bargaining possibilities. Let us then say that a strategy n-tuple is permissible if, in the psi possible coalition set, it is an equilibrium solution.

Let us say that it is permissibly permissible if it is such only in the psi set of this psi set. A strategy n-tuple will be obligatory if it is the maximum or unique equilibrium point in the psi set; it will be obligatorily obligatory if it remains such in the psi set of this psi set. In function of these definitions that can be formulated within the limits of one single game (this is the advantage of this proposal over the earlier one).

PPA does not imply PA, but PA does not imply PPA either; OOA implies certainly OA, and the reverse is not the case.

The two proposals each have their advantages; severe selection between them is not necessary; we shall have to pursue them both.

*d.* No act can be either obligatory or permitted without being possible, as only acts that belong to the set of A successors of the initial positions of the game can be members of equilibrium strategies. So OA implies Poss A and PA implies Poss A(in terms of the definition of possibility that we have adopted). It follows from these two identities that all contradictory acts are either indifferent or forbidden (because they cannot be permitted or obligatory, and because every act is either obligatory or prohibited or permitted or indifferent). But we can say equally that an act can never be the application of a strategy to a position if there are not other strategies that would apply other acts to that position. Id est: if A is either permitted or obligatory, not A is possible and so all necessary acts are either forbidden or indifferent.

Our interpretation gives clear indications as to the way in which this point (that von Wright discusses several times somewhat doubtfully) should be handled.

*e.* If an act is *strongly obligatory* (meaning: if it belongs to all strategies of the unique or maximal equilibrium n-tuple), an act that will apply to the same position a different operation or that will ap-

[4] See Luce et al., *op.cit*, pp. 245 and sq.

ply to the result of the first operation an operation reverting it, will not be present in any of the strategies of the equilibrium sequence (if not the two moves could not occur or could be deleted, and, by definition, the sequences are optimal) SOA implies thus that Not P Compl A (and F compl A). It seems also clear that if the complement of A is not permissible, it is forbidden and inversely.

What is said here for strong obligation does not however hold for weak obligation (if we have weak obligation for A, there could simultaneously be weak obligation for the complement).

*f*. We most emphatically do not have «OA or O compl A» (there will be games not having unique maximum equilibrium points).

We do not even have, for strong permissibility, such an analogue of the principle of excluded middle; indeed this sentence becomes in our interpretation the following one: A is part of all strategies in at least one equilibrium point, or there is an act B, incompatible with A or reverting A, and part of all strategies in at least one equilibrium point. This assertion is not true in general.

von Wright defends his principle of permission (wherein the distinction between strong and weak permission is naturally not made) as follows: if PA or P (compl A) were not true, not P (A) and not P (compl A) would be possible, but «not P(A) is F(A)», «not P(compl A) is F(compl A)». So F(A) and F(compl A) would be possible. This would imply that both not A and not not A would be obligatory, what is absurd. This chain of reasoning however uses two equivalences: FA is equivalent to O(compl A), and not P(A) is equivalent to F(A). We would defend these equivalences only as implications, in one direction.

Our dilemma then is the following one: should we look for a new interpretation of the complement of an act, that would make the principle true or should we look for changes in our P-O interpretation or should we to the contrary assert that here we possess a natural interpretation verifying most of the necessary P-O relationships, and proving, by not verifying the law under discussion, that intuition is not as consistent as we have thought ?

We do not claim to have a preference for one of these attitudes. We only want to point out the possibility of a choice.

*g*. We surely can assert that P(A or B) implies PA and PB and thus implies that A and B are not obligatory. The reverse is not true. To understand the acceptability of this assertion it is essential to presuppose that the acts are completely determined as to their place in the sequence of strategies; if they were only defined as species, the truth of these theses would not follow. But we have, from the beginning, defined acts by their initial positions, their finals and their operation.

To be more careful we might say: if A or B is strongly permitted, then A is weakly permitted or B is weakly permitted, or A is forbidden and B is strongly permitted, or B is forbidden and Á is strongly permitted.

*h.* The extensionality principle for P remains without noticeable change: if, whenever A happens, B happens, then, if P is permitted and only then, A is permitted.

*i.* Many discussions that occurred referred to the relations between permission, obligation and inclusion. It seems thus desirable that some remarks be made on this topic in the light of our interpretation.

Suppose that act A is strongly obligatory (and thus included in all strategies of the unique equilibrium point). Suppose that also act B is included in act A (B is the application of part of A's strategy to A's positions, or is the application of A's strategy to a part of A's position,) Suppose that this is obligatory (meaning: this inclusion is a fact in all the strategies of the equilibrium n-tuple) Then B must also be obligatory. However OA and «A includes only factually B« will not yield this result. Only OA and either O(A includes B) or OA and Nec(A includes B) will be sufficiently strong premises.

The main source of doubt in this whole discussion seems to be the following fact: if one is engaged in playing a non optimal strategy or attacking a non optimal strategy, it could very well be that the optimal strategy given these supplementary date (the inefficient opening moves of adversary or self) is not one of the strategies of the equilibrium point. The assertion that «A obligatory» and «(A includes B) obligatory» implies «B obligatory» holds only fully in the case of optimal plays.

*j.* The paradoxes of deontic implication constitute another of the sources of disagreement in this area. If A is obligatory (id est: in our simple minded version: the only equilibrium point), then if this optimal strategy is not selected, it is not at all the case that now any strategy n-tuple is optimal. There will be certain optimal strategies against non optimal ones, and they will not be arbitrary. No paradox will present itself.

*k.* Assertions like the following ones O(OA implies A) ask for special care in interpretation. Indeed O is an operator applying to acts, and OA implies A is not an act. However, let us interpret exceptionnally «OA implies A» as the name of the act consisting in making possible, real or necessary the connection expressed. This being the case our two theories of iterated modalities both make the assertion true.

*l.* Let us add that a logic of imperatives can also be developed within the limits of game theory (imperatives will be communications that persons privileged in the game structure use to direct the strategy of other players, depending upon them). In games with communications, the theory of imperatives is implicitly present.

We come to the following conclusion: interpreting acts, Boolean functions of acts, P and O as we did, we can derive a system of deontic modalities that is not an interpretation in the forma' sense of von Wright's system, but that deviates from this system only in cases where intuition would itself have doubts and that verifies many of the more important rules of the first deontic logic.

The fact that we can obtain such a large fragment of deontic logic on the basis of such extremely general interpretation confirms the impression that few, if any, among the specific features of moral reasoning are captured by the existent formalism, notwithstanding the wealth of known formulas.

To summarise the statements made before, let us try to clarify the position of the formal system for deontic modalities built upon this game theoretic interpretation, within the totaly of modal system Let it be clear that our system cannot lead to a member of Lewis' S-hierarchy. Iteration of modalities is naturally definable in our interpretation, and cannot be limited; thus we have neither S 5 (without iteration) nor S4 (with limited iteration). The two characteristic axioms of S5 and S4 were:

(a) The possible implies he necessarily possible,

(b) The possibly possible implies the possible.

Now it is not true that an act permissible in one of the games that are the outcomes of a permissible strategy of GI, is also permissible in GI. It is not true either that an act permissible in GI is permissible in all games that are outcomes of permissible strategies in GI.

The higher order modal operators are viewed in our interpretation as adverbs (id est: as modifying the extension of the first order modal operators). In a sense we thus can agree with von Wright who claims that no iteration is possible for deontic modalities stricto sensu (because they transform names of acts into sentences); but we disagree where he does not include in his list of possibilities the iterated deontic operator as a modifier. Two species of operators are thus used in this paper.

In order to investigate the relation between the first three S-systems and the ones discussed here, we should remember that we have a great multiplicity of modals in our interpretation. An act can be weakly permitted, or weakly strongly, or strongly strongly or strongly weakly (see definition page 76). Let us ask for the weakest and the strongest concepts of the series, in what relations they stand to the first three S systems.

The difference beween these two extremes is very large: it is the difference between an act present in one of the strategies of one n-tuple, and an act present in all strategies of all optimum n-tuples.

Notwithstanding this difference however, none of these operators satisfies the modal axiom «p strictly implies the permissibility of p» (axiom that would be the natural counterpart of «p implies the possibility of p»). But the two axioms characterising S2 and S3 are verified:

If one act includes another one, then, if the first is part of one strategy of one equilibrium n-tuple, the second is too; and if the simultaneous performance of two acts is part of one strategy of one equilibrium n-tuple, then also the performance of one of both is part of such a strategy.

A formal system, including propositional logic, having analogues of the S2 and S3 axioms and a selection among the sentences we have verified in our interpretation, as axioms, without having analogues for the S4, S5 or SI axioms: this seems to be the formalism, not yet completely described that should most adequately express the properties of our model.

But we shall only have reached the real uselfullness of our model when we begin to exploit the formal properties of the solution or core or equilibrium concepts in theory of games with the aim of developing the structure of deontic logic.

The mutual relationship of our different forms of permissibility become very clear if we remember that they stand to each other as the prefixes (Ex) (Ey), (Ex) (y), (x) (Ey), (x,y): SS implies WW, SS implies too SW and WS, SW and WS both imply WW; but SW and WS do not imply each other.

We must now ask two questions that shall be the subject of the second section of this paper: a) given the nature of ethical concepts are there any intrinsic reason to put into close contact game theory and the formalisation of a part of ethics that is deontic logic?

b) and, given an affirmative answer to this first question, are there in present day ethical theories already attempts that give sufficient information about the structure of the doubtlessly very peculiar game that we shall have to use, in order to really reach the ethical situation and that could furnish us somewhat more adequate game theoretical correlates of permission and obligation than the ones already studied here ?

## Section II

R. C. Braithwaite [5] has defended forcefully the point of view that solutions optimal in the game theoretical sense are ethical fair solution. His point of view has been attacked by many; recently however, more constructive criticism has been forthcoming [6].

Let us take as starting point that an ethical rule must order a series of preferences, persons and groups in a hierarchical order that is in some sense 'natural'. How can it be that the subordination of one interest under another is a natural one ? A desire has only one essential property: its wanting satisfaction. A subordination can thus only be natural if, through it, all or some relevant desires obtain more complete satisfaction than they would have obtained through mutual battle. The same holds for persons as wholes and groups as wholes. A multiplicity of interests, groups and persons is a multiplicity of players in a complex game that allows possibility of cooperation and coalition. The field of non zero sum n person games is the basis on whose foundations ethics should rest.

We could even say that the unrealistic assumptions of game theory (complete knowledge of possible actions, of possible outcomes, completeness of the preference relation, invariance of these features under deliberation) are, in its ethical application, less dangerous than anywhere else ! Among and before all duties the first is the duty to make our conflicts of duties conform to this ideal pattern, where the nature of the conflicting interests is clearly delineated, and where, alone, their natural subordination relations can be discovered.

[5] R. C. BRAITHWAITE *The Theory of Games as a Tool for the moral Philosopher* (Cambridge University Press, 1955).

[6] J. R. LUCAS, *Moralists and Gamesmen* (Philosophy, January 1959, pp. 1-12), JOHN RAWLS (Philosophical Review, v. 68, 1958, p. 177 and p. 176 of his article *Justice as Fairness)*.

LEO APOSTEL, *Cybernetika en Speltheorie als Hulpmiddelen der Ethika* (*Diogenes*, 1959).

But, unhappily enough, the theory of solutions for cooperative games, especially for n person cooperative games, is not at all as developed as the theory of two persons zero sum games. We can thus be certain that much research has to be done before we can hope, from this point of view, to derive concrete solutions for real ethical problems.

In order to point out, as much the affinity between the two problem situations, as the undecided character of this part of game theory, we want to point towards some common features of ethical and game theoretical research.

*a.* If an arbitration must be sought between n players willing to cooperate, it must certainly give to each player as much as this player would win alone (the equilibrium point of the isolated player), and, if possible, more.

*b.* The arbitration must not be such that there is at least one player in a coalition whose situation can be improved without harming any of the other members: this is called the demand for Pareto optimality.

*c.* The arbitration must simultaneously reflect the aid each player can give to the association by entering it and also the harm his departure would inflict on that same association. These two factors are however completely independent, and the strength of dependence of the arbitration function on both is problematic.

*d.* The arbitration must be as stable as possible; if two games have only slightly different outcomes, the arbitration must be closely similar.

These four rules are certainly as much implied by any ethical deliberation as necessary requisites for satisfactory solutions of n persou non zero sum games. The affinity is clear: account must be taken of all interests, positive account must be taken of all interests, and the degree of consideration given to each interest must depend on what the same can contribute or destroy.

As in Kenneth Arrow's 'aggregation problem', these same requirements are imposed on any satisfactory social preference function to be built upon the individual preference functions, we can say that the affinity between these three problems: the arbitration problem in game theory, the aggregation problem in welfare economics and the ethical problem must be one of the primary facts of future ethics.

*e.* The very uncertainties of the two theories are the same: nobody can conclusively assert that the possibility of choice is in itself valuable or valueless, or that the equalization of outcomes, or of sacrifices is to be considered an independent requirement for all arbitration functions. These are the very uncertainties of ethics.

*e.* Finally, must we make the preferences of the multitude of players comparable, even co-measurable or must we leave them in principle incomparable ? Braithwaite's solution to the moral problem is mainly the proposal of a certain technique of comparison (setting utility interval between minimal and maximum strategies for all players equal) accompanied by an arbitration in function of the threat advantage. Raiffa has proposed other techniques of measurement (relative advantage will be the guiding rule of the opponents). Nash has examined the case in which intercomparability is excluded. These particular solutions, advantage of game theory, that classical moral philosophy did not have, allow us to judge through their consequences the validity of the proposals. But the rejection of one of these still extrememely special proposals is by no means the rejection of the game theoretical approch to ethics. In fact Lucas has well recognised that the rejection of one of the solutions as an immoral ethics shows conclusively that game theory is capable of asking the ethical question. We only need to remain on a sufficient high level of generality not to confound game theory with a solution that has no privilege in its own domain.

We come to the conclusion that both on the basis of their certainties and of their uncertainties game theory and ethics are related.

*f.* But, we can say more: not only the general problem of n person non zero sum games is close to the ethical problem in its complete generality, but also the theory of a very specific type of game, that for facility reasons we shall call 'the ethical game'.

This ethical game has several interesting characteristics

1. If we consider needs, persons, groups and norms all as players, the ethical game is the game of being simultaneously engaged in a multiplicity of different games and finding outcomes, players, preference scales of these different games partially independent and yet function of one another.

In this respect: only arbitration rules for game mixtures can be compared to ethical rules. This is the reason why Boolean functions defined on games as wholes seem to have high potential usefulness. The solution theory of such games, though not completely unstudied, is yet very undeveloped.

2. If we consider sequences of games in such a fashion that our acceptance of a future game depends upon the results of an earlier one, we can reach a dynamical theory of games, that will allow preference ranking, strategy possibilities and other aspects of games to

change as the result of the playing itself. The ethical game is the game having as moves modification of its own structure. Luce and Raiffa in their remarkable work stress repeatedly that dynamic game theory has not yet been developed, but that sequential theory leads towards it.

3. Finally, both H.N. Castañeda and J. Rawls [7] consider ethical rules as highest order norms, norms about the way in which to organize all other norms. John Rawls makes this point even in the game theoretical context, saying that the ethical game is the game having as choices the different possible arbitration rules one might adopt for all future games. Ethical rules become thus universal arbitration rules. Adding our three last remarks together we come to the conclusion that ethical rules are rules for a) complex games b) dynamic games and c) universal games. To these three features we must join (as a conclusion from our earlier considerations) our view of ethical rules as arbitration rules for non zero sum, n person games, satisfying certain requirements of stability and symmetry mentioned before.

We think that these considerations furnish sufficient motivation to justify the first section of this paper *that wanted to connect the only existent attempt to formalise ethical reasoning: deontic logic and the only clearly described model of the ethical situation: game theory.*

However, the distance between the simple concepts of equilibrium points, used in our first section, and the complex theory of solution of mixed, dynamic and universal games, seems enormous. The reader might have the impression that after all we did not reach our aim: we connected deontic logic with one type of game theory, and the ethical situation with another. Who can be certain that these two types themselves are sufficiently related with each other ?

We cannot finish this paper before having shown at least the possibility of the bridge that is so much needed.

Let us first continue to relate permission to the solutions of a game, and obligation to the preferability of certain solutions. But let us modify the concept of solution itself (as we should do if we want to introduce possibility of communication, and mutual compensation).

Definition I. A strategy n-tuple is permitted if it leads to an imputation, and obligatory if its imputation belongs to the core.

[7] Castañeda *A Theory of Morality* (p. 339-352), Philosophy and Phenomenological Research, 1956 and Rawls, *op.cit.*

Definition 2. A strategy n-tuple is permitted, if it leads to an imputation, and obligatory if this imputation is in a solution.
We must explain the technical meaning of these terms.

Let S be a coalition of players (id est: a set of players determining together their strategy). Let S play against the coalition of all players not belonging to it. For this two person zero sum game, a value exists, corresponding to the equilibrium point. The value v computed for every possible coalition S, is called the characteristic function of members of coalitions, satisfying the following properties:

If i is a member of the coalition, the value for i is smaller than or equal to the outcome for i in the coalition b, the value for the coalition must not be smaller than the sum of the values for the individual members. These two requirements are the requirements of individual and group rationality.

We say that a sequence S belongs to the *core* if it is an imputation, and if moreover for any subset of S, the sum of values of its members is smaller than or equal to the value of this subset (we could call this subgroup rationality).

An imputation Y is said to dominate another imputation X with respect to a coalition T if and only if the value of Y is at least as great as the sum of the elements of Y and if for every member of the coalition T, $y_i$ is greater than $x_i$. A *solution* is a set of imputations that do not dominate each other, and that for each imputation outside of the solution includes an imputation dominating it.

We have now defined the key concepts present in our new version of the definition for the deontic operators.

We might propose more liberalised versions of these new definitions by not asking that solutions contain only imputations (and if we want to introduce moral values of sacrifice and renouncement this will be necessary presumably; even though we know that, by considering the ethical game as a mixed game, renouncement in one of the constituent games might even fall under the requirement of individual, group and subgroup rationality); but most probably it will be desirable to consider even more restricted versions. The following definition seems to have many advantages :

Definition III. A strategy is permitted if it leads to a solution, and is obligatory if it leads to a strong solution. Let us first explain the idea of a strong solution: outside of the sets of solutions by imputations dominating certain members of the set of solutions. Let us, starting with a given coalition structure, suppose that one or more members of the coalition, seduced by the attraction of such a dominating non-solution imputation provoke its adoption. Then as in the solution there must be an imputation dominating this imputation outside of it. Let us again suppose that some member of the game induces this corrected imputation. We shall now say that a solution is strong if for any such corrected deviation the resulting situation is for at least some of the deviants worse than it was in the beginning.

This concept, proposed by Vickrey (see Luce, 2) seems to grasp at least some essential properties of the norm, defining, if we can believe George Homans [8], a stability position of the human group.

Before now sketching some remarks about deontic modalities thus defined, let us point out that we introduce dynamic elements in the idea itself of the strong solution, and that the different requirements defining the imputation concepts seem to be concerned with several games simultaneously the coalition is engaged. The distance between our simplest proposal and our most adequate one is not yet crossed but we are at least walking in the right direction. If we can show that here too some theory of deontic modalities could be developed, we have every reason to hope that, once the future development of dynamic, mixed and universal game theory becoming a fact, on this basis a theory of deontic modalities can be grafted.

We can define Boolean functions for the sequences that are imputations, and for the classes of sequences that are the solutions.

One capital difference must be noted : a sequence is a relation. We shall thus have to use the relational operators. These Boolean or relational operators are even farther away from the Boolean functions of acts than are the Boolean functions of strategies. For the sake of simplicity a first study could start with the following supposition : an f function of imputations, is the imputation of the corresponding f function on strategies, while these strategies are themselves composed of f functions of acts, in their given order. Such an assumption has to be made if we are to relate the conditions on Boolean functions of imputations or solutions to those of acts.

We might also (to avoid some very paradoxical consequences) modify our definitions I and II in the following sense.

Definitions I', II', : is permissible any act belonging to a strategy leading to either a core or a solution, and is obligatory any act belonging to all strategies that lead to the core, or to all elements of one or of all solutions (there are many non identical solutions, all of them having many imputations as members).

We might even introduce the concept of an obligatory solution (that would be unique, or in some other sense maximal), or of an obliga-

[8] George Homans, *The Human Group*, (Wiley, 1952).

tory imputation (that would be an imputation returning in all solutions).

Let i and j be two imputations. Their intersection is not necessarily an imputation (even of a subgame). The individual rationality postulate will remain valid, but, given the fact that we only have (in opposition to the core) a group and not a subgroup rationality postulate, the intersection might have a sum lower than its individual value sum, and so the second imputation postulate will not hold. If the intersection of two imputations is an imputation nothing tells us that either i and j will be one.

Inside the core however where all subgroups will satisfy the group rationality postulate, the intersection will still be a core member.

Tentatively we would thus conclude (because we did not consider all relevant properties) that in definition I if two acts are permitted, their intersection is not necessarily so, but if two acts are obligatory, their intersection is (coming to this conclusion we use the simplifying postulate introduced on page 27 par. 4).

If i and j are two imputations, we could define union as the set of imputations that have one of the values of i and j in corresponding places. This set will have many members. If this is our definition of union we shall be certain that in this set there will be at least one imputation, but also that there will be many members that are not imputations. If i and j are both in the core, the union of both will contain members that are in the core, and if the union is in the core, both i and j will be necessarily in the core. The simplifying postulate could transform these assertions in assertions on acts.

Let us now consider solutions. We can study unions and intersections of imputations in solutions, or unions and intersections of solutions themselves (we could also introduce complements of imputations, being imputations taking as values the difference between a standard and the given values, but the artificiality of this procedure is rather evident).

The union of two solutions is not necessarily a solution, because no element inside a solution is allowed to dominate another element of that solution and because an element outside of a solution is allowed to dominate an element inside of it, if only it is itself dominated by some member of the solution. It is evident that the loseness of the solution concept (that contains perhaps thousands and thousands of incompatible strategies) makes here our combinatorial study rather difficult.

If the union of two sets of imputations is a solution nothing allows me to say that the two sets in isolation are solutions.

Similar remarks can be supplied by the reader with reference to intersections.

But the very fact that we can come to these negative conclusions, shows us that the concepts of solution, core and imputation can be put into relationship with a deontic logic; we should now ask what supplementary assumptions would give to unions and intersections of imputations and solutions a more stable character. The reasons of their present instability being clearly seen from the analysis that precedes, these reasons can also be eliminated. The very definition III (the concept of a strong solution) is, it seems to us, an attempt in the right direction, and simultaneously it is a more adequate expression of the intuitive concept of norm. We should recommend a thorough study of the syntactical properties of the concept of a strong solution and of the deontic logic that could be constructed, following the model sketched in our first section, on this foundation.

We must now bring this paper to a close. It is our opinion that the very difficulties we have met in our last part are evidence in favour of the assertion that the gap between our first and second section can be overcome; that what prevents us from already achieving success is exactly the feature that prevents concepts like solutions and imputations to be considered complete answers to the game theoretical problems.

A relationship has been established. The concepts of «obligation» and «permission» have received an interpretation that is simultaneously intuitive and formal. We avoided a too close relationship with particular theories of disputable character; and we provided the beginning of a semantics for deontic logic. We cherish the hope that others might continue in the same direction [9].

*(University of Brussels, University of Ghent.)*

[9] Among the set of possible acts, we find acts of asserting and denying. If we can develop a general deontic logic for acts without specifications, this deontic logic could be applied also to acts of asserting or denying. We could say that p is a consequence of q, if and only if it is not permitted to assert q and to deny p (permission being interpreted by the solutions of games wherein acts of asserting and denying are present as moves). Game theory could thus become, not only the foundation of deontic logic, but even of deductive logic in a more general sense.