

John-Jules's Strategic Mind

Paul Harrenstein and Mehdi Dastani *

1 Introduction

There is no great lie in saying that John-Jules is an ardent advocate of the employment of social and psychological concepts in logics for his beloved agent systems. His enthusiasm for formal analyses of norms, emotions, intentions, commitments and desires is quite capable of carrying him away. Nothing of the sort, alas, is also true for game theory, as to its merits John-Jules remains largely oblivious. No matter with how much fervor John-Jules manipulates semantical objects like belief sets, valuations and models, strategic equilibria are most likely to make his forehead frown. Even the celebrated Nash equilibrium is no exception in this respect. This is a pity, because game theory provides an elegant framework that seems to chime in well with John-Jules's overall interests. We believe, moreover, that a profound interest in game theory may allow John-Jules to have his precious single agents interact and socialize.

In an effort to understand this odd discrepancy, we venture the hypothesis that John-Jules has a logical rather than a strategic mindset. In this context, we observe that there is a fundamental difference between the strategy profiles of a strategic game and its possible outcomes. A strategy profile is a combination of strategies, for each player of the game one, whereas its outcome is intuitively a state of affairs that comes about as the result of playing the strategy profile. The latter are a bit like the possible worlds of Kripke semantics for modal logics and pretty much devoid of strategic content. In case our hypothesis proves to be correct, a rendition of game theoretic concept involving outcomes rather than strategies may be the way to arouse John-Jules's interest in the subject.

Nash equilibrium is a game theoretic solution concept which definition typically pertains to strategy profiles. A strategy profile is a Nash equilibrium whenever no player would benefit by unilaterally deviating from it. As such it gives content to a notion of strategic stability. In an equivalent fixed point formulation,

*The authors thank Jelle Gerbrandy for his useful comments on an earlier version of this paper.

a Nash equilibrium can be understood as combining strategies each of which is a best response with respect to the others.

The two aspects of a game that would thus seem to be relevant for singling out its Nash equilibria are the players' preferences and what each player can achieve by choosing a strategy or by deviating from a strategy profile. Usually, the preferences of the players are defined over the possible outcomes of the game and as such they can straightforwardly be represented by sets of subsets of outcomes. Moreover, it is possible to specify conditions for sets of subsets of outcomes that guarantee there to be a game of which the preferences can be represented thus. In a similar way, the manipulative powers of the players in a game determine unequivocally a family of sets of subsets of outcomes. With each player is associated the set of outcomes he can guarantee the game to end in by choosing an appropriate strategy. This set of subsets of outcomes we will refer to as the player's *effectivity set*. In his dissertation (Pauly (2001)), Pauly has successfully addressed the issue of which set theoretic conditions a set of subsets of outcomes has to satisfy if it are to constitute a family of effectivity sets for the players of a game. Moreover, he achieved a much similar result for an analogous notion of coalitional effectivity sets, i.e., the sets of subsets of outcomes a group of players can guarantee a game to end in by choosing appropriate strategies. Thus, to appraise whether a family of sets of subsets of outcomes corresponds to a family of effectivity sets, no *strategic* notions have to be invoked as such.

We find, however, that it is not in general possible to find the outcomes the Nash equilibria of a game give rise to on the basis of the players' preferences and their individual, or even their coalitional, effectivity sets alone. This observation seems to reconfirm Nash equilibrium as an essentially strategic concept and at the same time — assuming the validity of our hypothesis — it would explain John-Jules's contrariety to it.

Finally, we suggest, by way of experiment, an alternative equilibrium concept, *outcome equilibrium*, which does apply to outcomes rather than to strategy profiles. The outcome equilibria can be singled out on the basis of the players' preferences and their coalitional effectivity sets alone. No such result is possible if only the individual effectivity sets are taken into account. From the outcome of this experiment we draw our conclusions as to our hypothesis concerning John-Jules disinclination to think strategically. If he happens to like the concept, we consider our hypothesis to be corroborated. Nash equilibrium being about as elegant a game theoretic solution concept as you can wave a stick at, we find ourselves justified to attribute John-Jules's discord with it to its essentially strategic nature, a quality outcome equilibrium purposely lacks. In the other case, things are slightly more complicated. Were John-Jules to disapprove of outcome equilibria, it is not immediately clear whether this is due to him being disinclined towards strategic notions

or to our inability to formulate a likeable enough concept. Still, this paper, appearing as it does in his *liber amicorum*, is written in John-Jules's honor, and the reader may rest assured that every effort has been taken to render outcome equilibrium as appealing to John-Jules as possible.

2 Preliminaries

In this section we review some of elementary definitions of game theory as well as Pauly's results concerning the characterization of effectivity sets.

We define a *strategic game* as a tuple $(N, O, \{S_i\}_{i \in N}, \{\leq_i\}_{i \in N}, f)$, where N is a finite set of players, O is a finite set of outcomes,¹ and, for each player i , S_i is a countable set of strategies and \leq_i is a reflexive, transitive and connected relation, or weak order, over O . These latter relations reflect the players' preferences with regard to the outcomes. The set of *strategy profiles*, denoted by S , is given by the Cartesian product $\prod_{i \in N} S_i$. Finally, f is a function associating each strategy profile with an outcome in O . For technical convenience, we confine our attention to games with two or more players.

We also adopt the following notational convention. For C a subset of players and s and s' strategy profiles, we have (s_{-C}, s'_C) denote the strategy profile s'' such that, for all players i , s''_i equals s'_i , if i is in C , and s''_i equals s_i , otherwise. I.e.,

$$s''_i =_{df.} \begin{cases} s'_i & \text{if } i \in C, \\ s_i & \text{otherwise.} \end{cases}$$

In case C is a singleton $\{i\}$, we generally omit the parentheses. Accordingly, (s_{-i}, s'_i) denotes the strategy profile $(s_0, \dots, s_{i-1}, s'_i, s_{i+1}, \dots, s_n)$, where s is the strategy profile given by (s_0, \dots, s_n) .

A strategy profile s is said to *contain a best response* for a player i whenever $f((s_{-i}, s'_i)) \leq_i f(s)$, for all strategy profiles s' . A *Nash equilibrium* is then defined as a strategy profile that contains a best response for all players of the game.

Following Pauly (2001), we represent the (individual) *power*, or *effectivity*, of a player by a set of sets of outcomes E_i^G , also called the player's (*individual*) *effectivity set*. Each element of a player i 's effectivity set is a set of outcomes i can force the game to end in by choosing an appropriate strategy, no matter which strategies are chosen by his opponents. I.e., for each subset X of outcomes we have:

$$X \in E_i^G \quad \text{iff} \quad \text{there is an } s \text{ in } S \text{ such that for all } s' \text{ in } S: f((s'_{-i}, s_i)) \in X.$$

¹The assumption that the set of outcomes be finite is essential for Proposition 4.1, below.

We say that a family $\{\mathbf{X}_i\}_{i \in N}$ of sets of subsets of outcomes *corresponds to the (individual) powers of the players in G* , if \mathbf{X}_i equals \mathbf{E}_i^G , for each player i in N . Having assumed that each game involves at least two players, we have the following fact. For a proof, the reader be referred to (Pauly, 2001, page 22).

Fact 2.1 (Pauly) *Let $\{\mathbf{X}_i\}_{i \in N}$ be a family of sets of subsets of outcomes. Then there is a game G such that $\{\mathbf{X}_i\}_{i \in N}$ corresponds to the individual powers of the players in G if and only if, for each player i , \mathbf{X}_i is closed under supersets does not contain the empty set as an element, and $\bigcap_{i \in N} \mathbf{X}_i \neq \emptyset$, for any family $\{X_i\}_{i \in N}$ such that $X_i \in \mathbf{X}_i$, for each i in N .*

The concept of effectivity can straightforwardly and conservatively be extended as to apply to groups of players just as well as to individuals. Thus, each group of players C of a game G is associated with a set of subsets of outcomes, denoted by \mathbf{E}_C^G . For each group of players C , each element of \mathbf{E}_C^G is a set of outcomes the players in C can force the game to terminate in by choosing suitable strategies independently of the choice of strategy made by the players outside the group. Thus, formally, for each set X of outcomes and each subset of players C , we have:

$$X \in \mathbf{E}_C^G \quad \text{iff} \quad \text{there is an } s \text{ in } S \text{ such that for all } s' \text{ in } S: f((s'_{-C}, s_C)) \in X.$$

In particular, for $-i$ the group of all players except i , then:

$$X \in \mathbf{E}_{-i}^G \quad \text{iff} \quad \text{there is an } s \text{ in } S \text{ such that for all } s' \text{ in } S: f((s_{-i}, s'_i)) \in X.$$

Also observe that, in general, $\mathbf{E}_{\{i\}}^G$ equals \mathbf{E}_i^G . We say that a family $\{\mathbf{X}_C\}_{C \subseteq N}$ of sets of subsets of outcomes *corresponds to the coalitional powers in G* if \mathbf{X}_C equals \mathbf{E}_C^G , for each group of players C of N .

Pauly formulates a number of properties that guarantee a family of sets of subsets of outcomes $\{\mathbf{X}_C\}_{C \subseteq N}$ to correspond to the coalitional powers in a game. A family $\{\mathbf{X}_C\}_{C \subseteq N}$ is said to be *N -maximal*, whenever for all subsets X of outcomes, $-X \notin \mathbf{X}_\emptyset$ implies $X \in \mathbf{X}_N$. Furthermore, a family $\{\mathbf{X}_C\}_{C \subseteq N}$ is understood to be *superadditive* just in case for any *disjoint* groups of players C and C' , $X \cap X' \in \mathbf{X}_{C \cup C'}$, for all $X \in \mathbf{X}_C$ and all $X' \in \mathbf{X}_{C'}$. Intuitively superadditivity captures the idea that groups can coordinate their choice of strategy and thus achieve possibly more than the groups can separately. The following fact then holds. Again, Pauly's dissertation may be consulted for a proof (Pauly, 2001, Theorem 2.27 on page 28).

Fact 2.2 (Pauly) *Let $\{\mathbf{X}_C\}_{C \subseteq N}$ be a family of sets of subsets of outcomes. Then, $\{\mathbf{X}_C\}_{C \subseteq N}$ corresponds to the coalitional powers in a strategic game if and only if*

$\{X_C\}_{C \subseteq N}$ is N -maximal, superadditive and, for each subset C of N , the set X_C contains the full set of outcomes O but not the empty set and is moreover closed under supersets.

A similar, though much more commonplace, result can be obtained for the preferences of the players. The preferences of each player i can straightforwardly and unequivocally be represented by a set of subsets of outcomes by defining:

$$P_i^G =_{df.} \{\{o' : o \leq_i o'\} : o \in O\}.$$

The set P_i^G we will also refer to as *player i 's preference set*. Then for all outcomes o and o' we have:

$$o \leq_i o' \quad \text{iff} \quad \text{for all } Y \in P_i^G : o \in Y \text{ implies } o' \in Y.$$

A family of sets of subsets of outcomes $\{Y_i\}_{i \in N}$ is said to *correspond to the players' preferences in a game G* , whenever for each player i , the set X_i coincides with P_i^G . The following result can then easily be verified, it being given that the set of outcomes is finite.

Fact 2.3 *A set of subsets of outcomes $\{Y_i\}_{i \in N}$ correspond to the players' preferences in some game G if and only if, for each player i , Y_i is a chain in 2^O , ordered by the set inclusion relation \subseteq , which, moreover, contains as an element the full set of outcomes O but not the empty set.*

3 Outcomes and Nash Equilibria

The set of Nash equilibria of a strategic game are those strategy profiles that contain a best response for each player. Nevertheless, it is perfectly possible for an outcome that it is not yielded by any Nash equilibrium of a game, even though for each player i , it is the outcome of some strategy profile containing a best response for i . I.e., it is not in general the case that the subset of outcomes of a strategic game G given by:

$$\bigcap_{i \in N} \{f(s) : s \text{ contains a best response for } i \text{ in } G\}$$

equals the subset of outcomes given by:

$$\{f(s) : s \text{ is a Nash equilibrium in } G\}.$$

This phenomenon is illustrated by the two-player game $G1$ depicted in Figure 1. The one player, *Row*, chooses rows, whereas the other player, *Col*, chooses columns,

$$\begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

Figure 1. The two-player game $G1$ in which the intersection of the outcomes containing a best response does not coincide with the set of outcomes yielded by the Nash equilibria. Assume that for i either of the players, $a <_i b <_i c$. The only Nash equilibrium outcome, c , is in boldface.

determining one of the entries of the matrix as the outcome of the game. Let us assume that the both players prefer c to outcome b and value a least of all. The outcomes associated with the strategy profiles containing a best response for *Row*, as well as those containing a best response for *Col*, are then given by the set $\{b, c\}$. Hence, the intersection of the sets of outcomes yielded by a strategy profile containing a best response for one of the players coincides with $\{b, c\}$ as well. The only Nash equilibrium of this game is the strategy profile in which *Row* chooses the bottom row and *Col* opts for the right column. This strategy profile, however, yields c as the unique outcome.

This example, of course, merely shows that outcomes and strategy profiles belong to different categories. Two strategy profiles may have very different strategic properties and still give rise to the same outcome in a game. In a sense outcomes could thus be said to abstract from the strategic features of the underlying strategy profiles. The above example demonstrates, moreover, that by focussing on outcomes these strategic aspects of strategy profiles may be lost beyond recall.

The effectivity sets as they were introduced in the previous section are composed of sets of subsets of outcomes. A similar remark holds for the preference relations of the players in a game. The question we will be concerned with at this point is whether the structure of a game G that is preserved in a family $\{E_i^G\}_{i \in N}$ together with $\{P_i^G\}_{i \in N}$ suffices to single out precisely the set of outcomes yielded by the Nash equilibria of G . Or, to phrase the issue slightly differently, consider families $\{X_i\}_{i \in N}$ and $\{Y_i\}_{i \in N}$ which correspond to, respectively, the individual powers and the individual preferences of the players of a game G . Is it now possible to determine the set of Nash equilibrium outcomes of the game G ?

We find that the answer to these questions is to be negative. To appreciate this, consider the two-person game $G2$ depicted in Figure 2, where the players and their preferences are as in the game $G1$ of the previous example, i.e., *Row* chooses rows, *Col* chooses columns and both players prefer c to both a and b , and b to a . Consequently, we find that $\{P_i^{G1}\}_{i \in N}$ equals $\{P_i^{G2}\}_{i \in N}$. Moreover, the individual effectivity sets of both players in $G2$ coincide with one another just as

$$\begin{pmatrix} a & \mathbf{b} & b \\ b & b & \mathbf{c} \\ a & a & b \end{pmatrix}$$

Figure 2. The two-player game $G2$. Assuming the players' preferences as in $G1$, the two Nash equilibrium outcomes, c as well as b , are in boldface.

well as with those in $G1$, i.e., we have:

$$E_{Row}^{G1} = E_{Col}^{G1} = E_{Row}^{G2} = E_{Col}^{G2} = \{ \{a, b\}, \{b, c\}, \{a, b, c\} \}.$$

Hence, the same families of effectivity and preference sets correspond to the individual powers and preferences in both $G1$ and $G2$. Observe, however, that the game $G2$ has two Nash equilibria, whereas $G1$ has only one. Moreover, one of these Nash equilibria — viz., the strategy profile in which *Row* chooses the top row and *Col* the middle column — gives rise to a different outcome (viz., b) than the one in $G1$ (viz., c). We may conclude that in this case it is impossible to single out the Nash equilibrium outcomes on the basis of the effectivity and preference sets only.

To close this section, observe that things would have been no better if we had considered the collective instead of the individual effectivity sets. In both $G1$ and $G2$ the effectivity of the grand coalition of both players is given by $2^{\{a,b,c\}} - \{\emptyset\}$ and that of the empty coalition by $\{ \{a, b, c\} \}$. Observe, moreover, that apart from the singleton coalitions there are no other coalitions possible in a two-player game. Hence, the same pair of games still presents a counter example.

4 Outcome Equilibrium

To counterbalance the negative results of the previous section, in this section we will introduce a solution concept for which the collective effectivity and preference sets do suffice to single out its instances. The fundamental idea is to replace the concept of a strategy containing a best response by a notion of a player's satisfaction with a particular outcome. Formally we say that an outcome o is *satisfactory for a player i* in a game G , whenever $f(s) \leq_i o$ for some strategy profile s that contains a best response for i in G . The intuition underlying this concept is that a player should be able to reconcile himself with any outcome that is no worse than his best effort could have achieved in the least favorable circumstances, i.e., if his opponents had been conspiring against him and he still had managed to avert per-

$$\begin{pmatrix} a & e & d \\ c & d & c \\ b & a & e \end{pmatrix}$$

Figure 3. The two-player game $G3$.

sonal disaster. An *outcome equilibrium* of a game G we then define as an outcome that is satisfactory for all players of G .

As an illustration of this concept, consider the two-player game $G3$ depicted in Figure 3. Let in this case the preferences of the players, Row and Col , be antagonistic. In particular we assume $a <_{Row} b <_{Row} c <_{Row} d <_{Row} e$. Accordingly, for Col we have $e <_{Col} d <_{Col} c <_{Col} b <_{Col} a$. The strategy profiles containing a best response for Row , as can easily be checked, result in either the outcome c or e , of which he prefers c least. Hence, all of the outcomes at least as desirable as c — viz., c , d and e — are satisfactory for Row . In a similar fashion it can be established that the outcomes satisfactory for Col are a , b and c . So, in this case, c is the only outcome equilibrium of the game $G3$. Observe in this context that if d were eventually to emerge as the outcome of the game, Row can be certain not to have played a best response strategy. However, since he could have been worse off had he played a best response against Col 's choosing the left column, d still qualifies as a satisfactory outcome for Row .

But how much better than Nash equilibrium does outcome equilibrium fare with respect to its characterization in terms of players' preferences and effectivity? We find that the answer to this should be negative in case only individual effectivity is taken into account, whereas a characterization is possible if the richer structure of coalitional effectivity may be invoked.

As to the first, negative, claim, consider the games $G4$ and $G5$ depicted in, respectively, Figure 4 and Figure 5. Both games involve three players, Row , Col , and Mat . The first two choose rows and columns, as before, and Mat chooses matrices. Observe that in both games the effectivity of all the players is identical,

$$\begin{pmatrix} a & b & c \\ c & a & b \\ b & c & a \end{pmatrix} \quad \begin{pmatrix} b & c & a \\ a & b & c \\ c & a & b \end{pmatrix} \quad \begin{pmatrix} c & a & b \\ b & c & a \\ a & b & c \end{pmatrix}$$

Figure 4. The three-player game $G4$, in which c is the best, b the second best and a the worst outcome for each of the players. Its only outcome equilibrium is then c .

$$\begin{pmatrix} c & b & a \\ a & b & c \\ c & b & a \end{pmatrix} \quad \begin{pmatrix} b & c & a \\ b & b & b \\ c & a & b \end{pmatrix} \quad \begin{pmatrix} c & a & b \\ a & b & c \\ a & b & c \end{pmatrix}$$

Figure 5. Another three player game, $G5$. The players' preferences are as in $G4$. Here, apart from c , also b is an outcome equilibrium.

viz.:

$$E_{Row}^{G4} = E_{Col}^{G4} = E_{Mat}^{G4} = E_{Row}^{G5} = E_{Col}^{G5} = E_{Mat}^{G5} = \{ \{a, b, c\} \}.$$

Let, moreover, the preferences of the players over the outcomes in both games coincide, each preferring c to b , with a being valued least of all. Then, c is the only outcome equilibrium in $G4$. However, in $G5$ not only c but also b qualifies as an outcome equilibrium.

By contrast, the following proposition bears out the positive side of the issue.

Proposition 4.1 *Let $\{X_C\}_{C \subseteq N}$ and $\{Y_i\}_{i \in N}$ be families of sets of subsets of outcomes which, respectively, correspond with the coalitional powers and the preferences of the players of a strategic game G . Then, the outcome equilibria of G are given by the set:*

$$\bigcap_{i \in N} \bigcup_{X \in X_{-i}} \bigcap \{X \cap Y : Y \in Y_i \text{ and } X \cap Y \neq \emptyset\}.$$

Proof: First consider an arbitrary outcome o and assume it to be an outcome equilibrium. Then, for all players o is satisfactory, i.e., for each player i there is a strategy profile s containing a best response for i with $f(s) \leq_i o$. Observe that $\{f((s_{-i}, s'_i)) : s' \in S\} \in X_{-i}$. Then, for all strategy profiles s'' in $\{(s_{-i}, s'_i) : s' \in S\}$, we have $f(s'') \leq_i f(s)$. Define:

$$X^* =_{df.} \{f((s_{-i}, s'_i)) : s' \in S\} \cup \{o\}.$$

With X_{-i} being closed under supersets, it follows that $X^* \in X_{-i}$. Now consider an arbitrary $Y \in Y_i$ such that $X^* \cap Y \neq \emptyset$. Hence, either $o \in Y$ or $f((s_{-i}, s'_i)) \in Y$, for some strategy profile s' . If the former, we are done. In the latter case, there is, by definition, some outcome o' in O such that $Y = \{o'' : o' \leq_i o''\}$. It follows that $o' \leq_i f((s_{-i}, s'_i))$. By transitivity of \leq_i , then, subsequently, $o' \leq_i f(s)$ and $o' \leq_i o$. Consequently, $o \in Y$ and, hence, $o \in X^* \cap Y$.

For the opposite direction, assume for some arbitrary outcome o that it be no outcome equilibrium. Then, there is some player i such that $o <_i f(s)$, for all

strategy profiles s containing a best response for i . Now consider an arbitrary $X \in \mathbf{X}_{-i}$. If o is no element of X the case is almost trivial, as $X \in \mathbf{X}_{-i}$ contains a non-empty set and O is in \mathbf{Y}_i . So, for the remainder of the proof we may assume that $o \in X$. Moreover, there is some strategy profile s such that $f((s_{-i}, s'_i)) \in X$, for all strategy profiles s' . I.e., the set $\{f((s_{-i}, s'_i)) : s' \in S\}$ is a subset of X . With the set of outcomes being assumed to be finite and the players' preferences being weakly ordered, some reflection reveals that the set $\{(s_{-i}, s'_i) : s' \in S\}$ contains a strategy profile s^* that is a best response for i . Hence, both $o <_i f(s^*)$ and $f(s^*) \in X$. Let $Y^* =_{df.} \{o'' : f(s^*) \leq o''\}$. Then, by definition, $Y^* \in \mathbf{Y}_i$ and $o \notin Y^*$. Accordingly, $o \notin X \cap Y^*$. Finally, observe that with $f(s^*) \in X$, also $X \cap Y^* \neq \emptyset$, which concludes the proof. \dashv

5 Conclusion

Many existing game theoretic solution concepts, such as Nash equilibrium, apply to strategy profiles rather than to outcomes. We have argued that the outcomes generated by the Nash equilibria of a game cannot in general be determined in terms of the outcomes the players can force the game to end in, i.e., their effectivity, and the player's preferences alone. Apparently, the solution concept relies too much on the strategic structure of games as given by the strategy profiles. Therefore, we have provided an alternative game theoretic solution concept, called outcome equilibrium, which can be defined in terms of the players' coalitional effectivity together with the players' preferences. The strategic content of this concept being largely dissipated, we trust that it will appeal to John-Jules's logical predilections.

However, outcome equilibrium shares a number of supposedly noxious properties with Nash equilibrium. An outcome equilibrium is not guaranteed to exist and even if it does exist, it is not generally unique. Moreover, it is not in general the case that, for s and s' strategy profiles and i a player, if $f(s)$ and $f(s')$ are outcome equilibria, so are $f((s_{-i}, s'_i))$ and $f((s'_{-i}, s_i))$. Also, outcome equilibria need not be *Pareto efficient* in the sense that o may be an outcome equilibrium with there still being available a strategy profile s such that each player of the game prefers $f(s)$ to o . In any such case, however, $f(s)$ will be an outcome equilibrium as well. The verification of these facts we leave as an exercise to the honoree.

Reference

Pauly, M. (2001), *Logic for Social Software*. Ph.D. thesis, Institute for Logic, Language and Information, Amsterdam.