

*Economics and Philosophy*, 17 (2001) 221–234 Copyright © Cambridge University Press

# ON THE PRINCIPLE OF COORDINATION

MAARTEN C. W. JANSSEN

*Erasmus University Rotterdam*

---

## 1. INTRODUCTION

On many occasions, individuals are able to coordinate their actions. The first empirical evidence to this effect has been described by Schelling (1960) in an informal experiment. His results were corroborated many years later by Mehta et al. (1994a,b) and Bacharach and Bernasconi (1997). From the point of view of mainstream game theory, the success of individuals in coordinating their actions is something of a mystery. If there are two or more strict Nash equilibria, mainstream game theory has no means of explaining why people tend to choose their part of one and the same equilibrium. Textbooks (see, e.g., Rasmusen, 1989 and Kreps, 1990) refer to the fact that players may use focal points (see Schelling (1960)) or social conventions (see Lewis (1969)). Both notions cannot easily be incorporated into mainstream game theory, however. The notion of social conventions has recently been extensively studied in the context of evolutionary game theory where a population of agents interacts with each other. The central focus of this paper, however, is on situations where a few players play a game only once and I study how they may coordinate their actions.

Since Gauthier (1975), many authors have attempted to explain the use of focal points in a framework that departs in one way or another from mainstream game theory. Crawford and Haller (1990) try to explain how rational individuals can use past play to learn to coordinate. Bacharach (1993), Sugden (1995) and Janssen (2000) try to explain how

This paper replaces an earlier paper with the title *Towards a Justification for the Principle of Coordination*. I thank Michael Bacharach and Hans-Jorgen Jacobson for discussion of it. One anonymous referee gave unusually thoughtful and stimulating comments. I am more than grateful for these comments on earlier versions of this paper.

individuals can use the labels of strategies to coordinate their actions.<sup>1</sup> All these papers use variations of the Principle of Coordination, a term introduced by Gauthier (1975).

To get an idea of the differences between the different versions let us first look at the terms in which the different authors themselves phrase the Principle. Gauthier (1975, p. 201) defines the Principle in the following terms: 'in a situation with one and only one outcome which is both optimal and a best equilibrium, if each person takes every person to be rational and to share a common conception of the situation, it is rational for each person to perform that action which has the best equilibrium as one of the possible outcomes'. Crawford and Haller (1990, p. 580) 'maintain the working hypothesis that players play an optimal . . . strategy combination', which is defined as a strategy combination that maximizes both players' repeated-game pay-offs.<sup>2</sup> Similarly, the technical notion Bacharach (1993, p. 266) employs has players choosing an equilibrium strategy combination that is not strictly pay-off dominated by another equilibrium strategy combination. Finally, Janssen (2000) uses a principle, which basically says that if there is a unique Pareto-efficient outcome, then rational players will choose their part of it.

The main difference between these alternative uses of the Principle of Coordination is whether or not it applies to games that have multiple pay-off equivalent Nash equilibria. Crawford and Haller (1990) and Bacharach (1993) interpret the Principle as saying that players can coordinate on one of the Pareto-undominated Nash equilibria. Gauthier (1975) and Janssen (2000), on the other hand, use the Principle more narrowly as applying only to games with a unique Pareto-efficient Nash equilibrium. Another difference relates to whether the Principle applies when none of the Nash equilibria of the game is Pareto-efficient. Bacharach's notion seems to select the most efficient equilibrium even if it is Pareto-dominated.<sup>3</sup> Gauthier, on the other hand, restricts the use of the Principle to cases where there is a unique Pareto-efficient outcome, which is then also a Nash equilibrium.<sup>4</sup>

The different versions of the Principle of Coordination are not uncontroversial. The Principle has been criticized by Gilbert (1989, 1990), among others. The purpose of this article is not to remove all controversy. Rather, the purpose is to clarify the discussion by showing that one version of the Principle follows from some axioms about considerations individual players have when choosing their actions. The discussion

<sup>1</sup> An overview of this literature is given in Janssen (1998a).

<sup>2</sup> The paper restricts attention to *attainable* strategies. As this concept is not important for my purposes, I have not mentioned it in the definition introduced in the main text.

<sup>3</sup> I say 'seems' as the paper does not discuss games where this is relevant.

<sup>4</sup> Sugden's principle is a little more subtle and will be discussed in Section 3.

about the use of the Principle can then proceed by arguing whether or not these axioms are intuitively appealing and whether or not we should abandon the variations of the Principle that cannot be derived from these axioms. The analysis below provides support for the Principle of Coordination only in so far as there is a unique, strict Pareto-efficient outcome and the basic idea of the rationalization is that each player individually forms a plan specifying how each player will play the game and which conjecture to hold about the opponent's play. The axioms that are postulated are at the level of these individual plans. We show that these axioms are such that if there is a unique strict Pareto-efficient outcome, then there is a unique plan how to play the game. As the plan is unique, both players *thinking individually* will play according to the same plan and the Pareto-efficient outcome results.

The problem of *coordination* is different from the problem of *cooperation* as exhibited, for example, in the prisoner's dilemma. As defection is the dominant strategy, one has to depart from the axiom of individual rationality in order to explain cooperation in the prisoner's dilemma. That is not the case for the problem of coordination: in order to explain that individual players will coordinate an approach consistent with the traditional axioms of individual rationality may be taken. The difference between the problem of *coordination* and the problem of *cooperation* is clearly recognized by Gauthier, for example. In his article *Coordination*, Gauthier (1975, p. 199) briefly mentions the prisoner's dilemma and the wish of agents to agree on a principle that yields an optimal outcome. He then says

insofar as they seek to do so, their problem is one, not of coordination but of cooperation – acting together to secure a mutual benefit unavailable to those who act independently to secure individual benefit. This problem is not our concern. We shall restrict discussion to situations where it is not advantageous or not possible to cooperate in attaining an optimal non-equilibrium outcome.

The Principle of Coordination that he sets forth, which is quoted at the beginning of this article, only applies to situations in which there is 'one and only one outcome which is both optimal and a best equilibrium'. Hence, this principle does not apply to prisoner's dilemma games which have more than one optimal (read: Pareto-efficient) outcome and Gauthier leaves a discussion of the problem of cooperation to his *Morals by Agreement* (1986). The present paper also restricts attention to coordination issues.

One approach to arguing in favour of the Principle of Coordination is based on the idea that when playing a (coordination) game individuals should regard themselves as members of a team. Team thinking has recently been studied by Sugden (1991, 2000) and Hollis (1998), among

others. Sugden (1991, p. 776) develops the notion of team rationality within the tradition of regarding game theoretic solution concepts as being formalized in a 'book of recommendations for playing games which is entirely authoritative', the recommendation being addressed to *both* players (rather than to each of the players individually). Sugden (2000, p. 183) analyzes coordination issues between football players as an example and introduces a coach making recommendations to the players on how to coordinate their play. Despite some similarities (to which I come later), the approach I take in this paper differs from the team rationality approach in two important ways. First, in many actual coordination problems there is no coach that is in the position to make recommendations to *both* players, or to use Sugden's earlier metaphor, there is no book of recommendations players can consult. Rather, *each* player *individually* has to develop a view on how both players will play the game and they individually have to look for arguments that support this view. The arguments individual players seek are formulated in the axioms briefly mentioned above and which will be developed in more detail below. In brief, the approach I take in this paper is that if the game structure is such that there is only one view satisfying the axioms, then the reasoning of individual players must result in the same view and, hence, players will coordinate. A second difference is that the present approach is more in line with mainstream game theory in the sense that I require (as one of the axioms) a player's view on how to play the game to be consistent with individual rationality. Hence, according to each player's view each individual chooses an optimal action given the conjectures about the opponent's play. This implies that the approach advocated here cannot explain cooperation in the prisoner's dilemma.

The formal approach I take in this paper is related to an approach advocated by Jacobson (1996) in this journal. Jacobson (1996) investigates an alternative eductive foundation for the notion of Nash equilibrium. He argues that players individually formulate a plan how a game will be played. An individual plan specifies for each player a set of pure strategies and a set of conjectures about the opponents' play. The formulation of a plan incorporates the game theoretic idea that each player not only thinks about what he himself is going to play, but also imagines himself in the position of his opponent. In the present paper I also take as a starting point the idea that individual players formulate a view (plan) about how both players play the game and which conjectures they have about the opponent's play. Like Jacobson I require a plan to be rational, but instead of imposing a requirement of internal consistency I require a plan to be *optimal* in the sense that a player will only adopt a particular view (plan) if another rational plan that is strictly better for both players does not exist. I show that the uniqueness version of the Principle of Coordination follows from the two requirements (rationality

and optimality) taken together. Hence, the present paper can be considered as providing a justification for it.<sup>5</sup>

The approach towards the Principle of Coordination I want to develop in this paper is a middle ground between mainstream game theory with notions like individual preferences and individual rationality, and the team rationality approach developing notions like 'we-thinking' and 'team preferences'. Like the team rationality approach, I regard the fact that players play (a game) together as implying that each of the players realizes they have something in common, namely (at least) the game they play, and that they depend on each other. This realization on the part of the players materializes in two ways. First, each player formulates a plan about how, according to each of them, both players will play the game and which conjectures both players hold about each other's behaviour. Second, if a plan exists which is unambiguously in everyone's interest, then the individual players will not hesitate to choose their part of the plan. On the other hand, like traditional game theory, the approach advocated here builds on the notions of individual preferences and individual rationality. What is in a player's own interest is simply defined by her own utility function and players only consider plans that are consistent with standard individual rationality considerations.

The plan of the paper is as follows: Section 2 presents the main axioms; Section 3 gives a formal statement of the result; Section 4 provides a discussion; and Section 5 provides conclusions.

## 2. GAME STRUCTURE AND AXIOMS OF PLAY

The analysis that follows is built around the following two-player game. There is a Row player and a Column player. Each of the two players has a finite strategy space, denoted by  $S_R$ , respectively  $S_C$ , where  $S_R = \{1, \dots, m\}$  and  $S_C = \{1, \dots, n\}$ . The pay-off to player  $i$ ,  $i = R, C$  is given by  $\pi(r, c)$ , where  $r \in S_R$  and  $c \in S_C$ . The structure of the game, including the pay-offs is common knowledge.

In this section I introduce three axioms on which the analysis is built. The first two axioms are based on Jacobson (1996) and are quite uncontroversial from the point of view of mainstream game theoretic reasoning; the third axiom is new. The first axiom formulates the idea that players plan what to do in their own situation and also imagine

<sup>5</sup> Colman and Bacharach (1997) have recently provided another justification. They use a *Stackelberg heuristic* which is defined as a way in which the players can *conceive* the game as being played sequentially (whereas it is actually played simultaneously). A critique that can be levelled against their approach is that coordination games are really simultaneous move games and we know, from the standard game theoretic literature, that simultaneous move games are rather different from sequential games. Hence, it is difficult to explain why players would *conceive* of the simultaneous move game in sequential terms.

themselves in the situation of their opponent. In other words, when playing a game, players realize that they depend on each other and, therefore, they think about what the others may do and what they may believe about themselves. Formally, a plan specifies for each player a set of pure strategies that are motivated by the plan and a conjecture what the other player will choose.

**A1.** Each player formulates a plan  $P$  with  $P = (R, q; C, p)$ ,  $R \subseteq S_R$ ,  $C \subseteq S_C$ ,  $q \in \Delta_C$ , and  $p \in \Delta_R$ , where  $R$  and  $C$  are the sets<sup>6</sup> of pure strategies motivated by  $P$  for the Row and Column player, respectively, while  $q$  and  $p$  are the conjectures (about the opponent's play).

Not every plan is a reasonable plan. First, it seems reasonable to require that a plan be such that the sets of strategies that are motivated by the plan are best responses to the conjectures that are held about the other player's play. This incorporates the idea that every player realizes that none of them acts against their own interest. To this end, let us define by  $\bar{B}_C(p)$  and  $\bar{B}_R(q)$  the set of pure strategies that are a best response to  $p$ , respectively  $q$ . A2 formulates the first condition a reasonable plan has to fulfill.

**A2.** (Rationality). A plan  $P := (R, q; C, p)$  has to be such that the pure strategies motivated by  $P$  are best replies to the conjecture held about the opponent's play, i.e.,  $R \subseteq \bar{B}_R(q)$  and  $C \subseteq \bar{B}_C(p)$ .

It is clear (and convincingly argued by Gilbert (1990)) that the Principle of Coordination does not follow from individual rationality considerations alone. Hence, we have to be more restrictive and impose another requirement that a reasonable plan has to fulfill. To this end, let us note that  $\bar{B}_C(p)$  and  $\bar{B}_R(q)$  are finite sets so that for each plan  $P$  and for each player  $i$  the minimum and the maximum pay-off associated with that plan are well defined and we denote them by  $\underline{\pi}_i(P)$  and  $\bar{\pi}_i(P)$  respectively. The third axiom requires that players only adopt plans that are optimal. A plan  $P$  is considered to be an optimal plan if the maximum pay-off both players get if they follow this plan is larger than the minimum pay-off both players would get according to any alternative plan satisfying A2.<sup>7</sup> This is the content of A3.

**A3.** (Optimality). A player's plan  $P = (R, q; C, p)$  must be such that there does not exist another plan  $P' = (R', q'; C', p')$  satisfying A2 such that

- (i)  $\underline{\pi}_i(P') \geq \bar{\pi}_i(P)$  for  $i = 1, 2$  and
- (ii) for each player  $i$  if  $\underline{\pi}_i(P') = \bar{\pi}_i(P)$ , then  $\bar{\pi}_i(P') > \underline{\pi}_i(P)$ .

<sup>6</sup> Note that the set of pure strategies  $R$  and  $C$  that are motivated by a plan need not be singletons.

<sup>7</sup> Here, I take a conservative view on optimality. One could defend a stronger criterion, namely that a plan is reasonable if there does not exist another rational plan such that both players receive a pay-off that is at least as large and at least one player gets a strictly larger pay-off, i.e., condition (ii) below needs to hold for one player only.

Note that A3 does not involve any team preferences (Sugden, 2000) or 'we-rationality'. The axiom is completely stated in terms of individual preferences and can be considered to take place in an individual player's mind. One reason A3 may be a reasonable axiom to impose on individual plans is the following. Suppose that after the game has been played, there is a (fair) chance the two players will meet and a discussion ensue about the reasons the players had for their choices. Suppose, moreover, that one player, Olga, used the Optimality criterion A3 to guide her plan, and the other, Soren, chose a Suboptimal plan. Olga may start the discussion by saying (somewhat angrily): 'OK, Soren, I understand you want to get the best pay-off for yourself and don't care too much about the pay-off I get, and I also understand that there are many choices we could make so that there is no a priori reason we should coordinate our choices, but I don't understand why you sacrifice both my pay-off and your own for . . . I don't know what'. Soren, somewhat taken aback, may reply by saying: 'I'm awfully sorry, I was thinking too much about other considerations and failed to see that it is actually also in my own interest to act differently from what I had planned for both of us'. Moreover, before playing, the players may foresee such a dialogue taking place after they have played the game in which the player who followed the optimality criterion has reason to accuse the other of not forming an optimal plan, while the other will be forced to apologize for her behaviour. If players prefer not to be in a position where they have to apologize, they will choose to opt for the optimality axiom A3.<sup>8,9</sup>

### 3. RESULTS

In this section I briefly point out the main implications of requiring a reasonable plan to fulfill A1–A3. Before I do so, some terminology is introduced. I will call an outcome *strictly Pareto-efficient* if the pay-off to each of the players in this outcome is strictly larger than the pay-off to each of them in any other outcome. Moreover, two plans  $P$  and  $P'$  are *outcome equivalent* if they support the same strategies, i.e.,  $R=R'$  and  $C=C'$ .

<sup>8</sup> This argument may be framed in terms of a simple normal form game with opting and not opting for axiom A3 as the two possible choices. No matter whether the other player opts for A3, it is better to opt for it if you prefer not to apologize. Hence, opting for A3 is a dominant strategy.

<sup>9</sup> The above argument bears some similarity to the argument put forward by Gilbert (1990, p. 16), where she argues that an interchange may lay the foundations for players to jointly accept a principle as 'our' principle. The main difference with the approach above is that the principle of optimality A3 is based on pay-off considerations and that people may dislike to apologize. This makes the above argument more in the spirit of individual rationality considerations than Gilbert's.

A first result says that a plan satisfying A1–A3 always exists. Formally, this result is trivial (and a proof is omitted), but it is nevertheless interesting to note that its existence marks an important difference from the approach taken by Sugden. Sugden (1995, p. 542) defines a recommendation  $A^*$  to be collectively rational if there exists a pair of utilities  $(u_1^*, u_2^*)$  such that if players choose strategies from  $A^*$ , they will get a utility outcome of  $(u_1^*, u_2^*)$  while each of them gets a strictly lower utility level if at least one of them chooses differently. He then proceeds by defining the principle of collective rationality: if a collectively rational recommendation exists, each player should act on that recommendation. It is clear, however, that a collectively rational recommendation does not always exist. To mark the contrast, I formally state the first result.

**Proposition 1.** For any game considered here, a plan  $P = (R, q; C, p)$  satisfying A1–A3 exists.

Proposition 2 below states the main result.

**Proposition 2.**<sup>10</sup>

- (i) If there exists a strictly Pareto-efficient outcome, then the set of plans satisfying A1–A3 is outcome equivalent and players will coordinate on the Pareto-efficient outcome.
- (ii) If the set of plans satisfying A1–A3 are outcome equivalent, then players coordinate on the most efficient Nash equilibrium.

*Proof.* (i) If there exists a strictly Pareto-efficient outcome, then there are  $\sigma$  and  $\rho$  such that  $\pi_i(\sigma, \rho) > \pi_i(\sigma', \rho')$  for  $i = 1, 2$  and for all  $\sigma'$  and  $\rho'$  with  $\sigma, \sigma' \in S_R$  and  $\rho, \rho' \in S_C$  and  $(\sigma, \rho) \neq (\sigma', \rho')$ . Hence,  $(\sigma, \rho)$  must be a Nash equilibrium. Accordingly, there are conjectures  $p$  and  $q$  such that  $P = (\sigma, p, \rho, q)$  satisfies A2. Moreover, all other plans that are not outcome equivalent are suboptimal and, hence, do not satisfy (A3). Thus, the set of plans satisfying A2 and A3 is outcome equivalent.

(ii) Suppose the set of plans satisfying A2 and A3 is outcome equivalent. This implies that there exists a plan  $P = (R, q; C, p)$  such that for all and  $p'$  and  $q'$  that are part of another plan  $P' = (R', q'; C', p')$  with  $R' \neq R$  or  $C' \neq C$  satisfying A2,  $\underline{\pi}_i(P) \geq \bar{\pi}_i(P')$  for  $i = 1, 2$  and for each player  $i$  if  $\underline{\pi}_i(P') = \bar{\pi}_i(P)$ , then  $\bar{\pi}_i(P') > \underline{\pi}_i(P)$ . The fact that all plans satisfying A1–A3 are outcome equivalent implies that  $R$  and  $C$  contain

<sup>10</sup> If a stronger version of the optimality axiom A3 is used, then we can assert a stronger version of this Proposition. In particular, we may then drop the ‘strictness’ requirement in part (i).



just one strategy. It is also clear that if a Nash equilibrium is selected, then it will be the most efficient one. Suppose, then, that on the contrary  $(R,C)$  does not constitute a Nash equilibrium. Without loss of generality assume that given the choice of the Column player, the Row player could improve his position, that is, there exists an  $R'$  such that  $\pi_R(R',C) > \pi_R(R,C)$ . Take  $R'$  to be (one of) the best response(s) to  $C$ . It is clear that  $(R',C)$  itself cannot be part of a plan satisfying A1–A3 as this would violate the fact that the set of plans satisfying A2 and A3 is outcome equivalent. So, suppose that there does not exist a plan that satisfies A1–A3 that supports  $(R',C)$ . As the plan supporting  $(R,C)$  satisfies A3 and as  $\pi_R(R',C) > \pi_R(R,C)$ , this would imply that no plan exists supporting  $(R',C)$  that satisfies A2. It is easy to see, however, that this also cannot be the case as (i) the plan supporting  $(R,C)$  satisfies A2 and so that certainly  $C$  can be supported as a rational choice and as (ii)  $R'$  can be supported as it is the rational choice given the conjecture that the other plays  $C$ . ■

Note that Proposition 2 does not imply a Pareto-efficient outcome result if axioms A1–A3 are satisfied. When there are multiple Pareto-efficient outcomes, none of them may be compatible with the rational choice of both players – as in the prisoner’s dilemma. Hence, the first part of Proposition 2 does not guarantee that a Pareto-efficient outcome results if players’ plans satisfy A1–A3. In particular, applied to the prisoner’s dilemma, axioms A1–A3 only support the Pareto-dominated Nash equilibrium.

Related to the different version of the Principle of Coordination, it is also important to note that for many games there may, in addition, be multiple plans satisfying A1–A3 which are not outcome equivalent. Hence, in general, it is not guaranteed that players coordinate their behaviour. Not only do axioms A1–A3 not select a set of plans that is outcome equivalent in case a game has multiple pay-off equivalent Nash equilibria, they may also fail to select between outcome non-equivalent plans in case the game has a Nash equilibrium that is better for both players than all other Nash equilibria, but is itself not Pareto-efficient. Recall from the Introduction, that Bacharach’s version seems to select the most efficient Nash equilibrium. In the example of Table 1, there is a plan that supports the outcome  $(U,L)$ , for example, one that specifies that players conjecture the other player to randomize with probability  $1/3$  over all actions.

In summary, the proposition provides a possible justification for the uniqueness version of the Principle of Coordination, but it does not provide support for the other versions mentioned in the Introduction.

	<i>L</i>	<i>M</i>	<i>R</i>
<i>U</i>	7,7	5,8	7,4
<i>C</i>	8,5	6,6	0,0
<i>R</i>	4,7	0,0	4,4

TABLE 1. An efficient Nash equilibrium relative to the set of all Nash equilibria may not be the unique plan satisfying A1–A3.

#### 4. DISCUSSION

I realize that the Principle of Coordination is a controversial principle. In this section I discuss some of the possible critiques along two lines: (i) coordination is assumed rather than explained; (ii) what is the relationship to risk dominance? The purpose of the discussion is to make the argument more transparent by describing what is entailed, and especially, what is not.

Before I do so, however, it is important to realize that even the weakest solution concepts that are employed in non-cooperative game theory, like iterative elimination of dominated strategies (IEDS) or rationalizability, have individual players impose restrictions on the likely behavior of others (see Bernheim, 1984 and Pearce, 1984). The common knowledge of rationality assumption on which IEDS is based assumes that all players conjecture (or even: know) other players to be as rational as they are. It is clear that this assumption imposes restrictions beyond the simple notion of rational individual behavior. Nevertheless, the assumption is well accepted in non-cooperative game theory, so much so that it is hard to encounter justifications for it. One way to interpret the assumption is that while thinking about what the other may choose, players impose their own kind of rationality on their opponents. According to this interpretation the assumption only imposes restrictions on the thought processes of individual players (see Janssen, 1998b). Hence, one may defend the thesis that the considerations that lead to the notion of IEDS are purely individualistic in nature and do not involve any notion of team thinking.

The requirement I impose in addition, namely, that plans be optimal, can be justified along similar lines. Each player formulates an individual plan. To be reasonable, a plan fulfills a requirement that goes beyond individual rationality. While thinking about what others may choose, players come to the conclusion that some actions may unambiguously be in their mutual interest. If a player of the game has to come up with a plan that stipulates what both players will do, he should formulate the

best possible plan taking into account that players act in their own interest. Any other plan hurts him as well as the other player and is difficult to support in a conversation that may follow after the game has been played. This consideration imposes restrictions on the reasoning process individual players go through before adopting a plan. Hence, axiom A3 can be justified as being of an individualistic nature.

### Coordination

One possible objection to the Principle of Coordination is that it assumes a form of coordination instead of explaining it in terms of individual considerations. The above analysis is, at least partly, able to counter this criticism. To this end, consider the game of pure coordination given below.

Axioms A1–A3 do not guarantee in any way that players coordinate in the game in Table 2. One player's plan may involve both players choosing  $L$  and conjecturing that the other will play  $L$ , whereas the second player's plan may specify that both players choose  $R$  and conjecture that the other will play  $R$ . As this form of miscoordination is not excluded by A1–A3, I conclude that the Principle of Coordination that follows from A1–A3 does not assume coordination.<sup>11</sup>

More generally, if there are multiple equilibria that cannot be Pareto-ranked, then A1–A3 do not guarantee that players coordinate their actions on a Nash equilibrium. This is important as some authors have assumed that players can coordinate on one of the Pareto-efficient equilibria in case of multiplicity. Bacharach (1993, p. 266), for example, assumes that players choose Pareto-undominated Nash equilibrium strategy combinations. A similar problem arises in the analysis of Crawford and Haller (1990, p. 575) who assume that players can maintain coordination in an infinitely repeated version of Table 2 if they have coordinated once. As there are potentially many different ways in which players can maintain coordination, and as each of these ways is equivalent in terms of pay-offs, our analysis does not provide foundations for this assumption (see Goyal and Janssen, 1996).

We can take this point further and argue that even in the case of a unique Nash equilibrium, our approach does not guarantee that players coordinate their choices on the equilibrium point. The example in Table 3 (a variation on a game due to Bernheim (1984)) makes this clear. Each player has three strategies from which to choose and each of these strategies is a rational choice given some conjecture. Moreover, there is no rational plan that is suboptimal. Hence, axioms A1–A3 do not impose restrictions on the possible choices players make. There is, however, only one Nash equilibrium, namely  $(C, M)$ .

The latter example also makes clear the difference between Jacobson

<sup>11</sup> A similar conclusion follows when considering a game like the battle of the sexes.

	<i>L</i>	<i>R</i>
<i>L</i>	1,1	0,0
<i>R</i>	0,0	1,1

TABLE 2. A game of pure coordination.

	<i>L</i>	<i>M</i>	<i>R</i>
<i>U</i>	0,7	2,5	7,0
<i>C</i>	5,2	3,3	5,2
<i>R</i>	7,0	2,5	0,7

TABLE 3. Reasonable plans may involve non-Nash behavior.

(1996) and the present paper. Jacobson requires plans to be internally consistent in the sense that players do not consider the possibility that the other player uses a different plan. This implicitly assumes that players *believe* they will coordinate. Requiring plans to be internally consistent, Jacobson (1996), shows that when there is a unique Nash equilibrium, individuals will coordinate on it. Applying his approach to the example in Table 3 implies that players will choose their part of the Nash equilibrium (*C,M*). My approach, in contrast, does not impose any restriction in that example. Indeed, if we have no grounds for assuming that players will coordinate, then we also have no grounds for assuming that players believe they will coordinate. Accordingly, there does not seem to be a good reason to impose the axiom of internal consistency on individual plans.

### Risk Dominance

A second possible objection to the Principle of Coordination may be that it does not hold when there is a conflict between Pareto-efficiency and risk dominance. An example is given in Table 4. In that figure there are two Nash equilibria in pure strategies: (*T,L*) is Pareto-efficient and (*B,R*) is risk dominant.<sup>12</sup>

Recent literature in game theory has resulted in conditions under which players are expected to play the risk dominant rather than the

<sup>12</sup> An equilibrium is risk dominant if both players' best response remains unchanged as long as the opponent chooses the equilibrium strategy with a probability at least equal to 0.5.

	<i>L</i>	<i>R</i>
<i>L</i>	10,10	4,7
<i>R</i>	7,4	8,8

TABLE 4. A conflict between Pareto-efficiency and risk dominance.

Pareto-efficient equilibrium. Carlsson and Van Damme (1993) consider a framework in which players observe pay-offs with some noise and show that (under some conditions) the risk dominant equilibrium survives IEDS. Kandori, Mailath and Rob (1993) and Young (1993), among others, study a population of agents interacting in an evolutionary environment and show that the risk dominant equilibrium is selected in the long run.

The circumstances I consider in this paper do not fit either one of these environments. This paper provides an *eductive* justification for the Principle of Coordination in case the game structure, including the pay-off, is common knowledge. In particular, the justification for the axiom of optimality (A3) I have developed at the end of Section 2 does not hold in these other environments. Two elements are important in this justification. First, there is a fair chance that agents can hold the other accountable for their actions when they meet after they have played the game. Second, the accountability for actions is based on common knowledge of pay-offs. Olga, at the end of Section 2, could argue: 'Listen, Soren, we both know the pay-offs for sure and we know that the other does. Hence, we know that the pay-offs each one of us would get under your plan are lower than the pay-offs we would get under my plan . . .'. The literature mentioned above considers other environments and I do not want to argue that the Principle of Coordination should apply in each and every possible situation. When there is no common knowledge about pay-offs (as in the Carlsson and Van Damme (1993) framework), or when there is almost no chance that she can be held accountable for my actions (as in evolutionary game theory), Olga does not have reasons to (expect to be able to) start a conversation in this way. Hence, the justification for A3 fails in these environments.

## 5. CONCLUSION

This paper has tried to clarify the Principle of Coordination by providing three axioms from which the Principle follows. By doing so, I have been able to tell what the Principle entails and what it does not, making the controversy around the Principle more transparent. For example, the Principle does not tell players in a prisoner's dilemma game to cooperate. Also, it does not solve the battle of the sexes, nor coordination

games where the Nash equilibria have identical pay-offs. In this way the paper discriminates between different versions of the Principle of Coordination. The axioms mentioned here only support a 'uniqueness' version of the Principle, which basically says that if a unique Pareto-efficient outcome exists in a game, then rational players will choose their part of that outcome.

#### REFERENCES

- Bacharach, M. 1993. 'Variable universe games'. In *Frontiers of Game Theory*. K. Binmore, A. Kirman and P. Tani (eds.). MIT Press
- Bacharach, M. and M. Bernasconi. 1997. 'The variable frame theory of focal points: an experimental study'. *Games and Economic Behavior*, 19:1–45
- Bernheim, D. B. 1984. 'Rationalizable strategic behavior'. *Econometrica*, 52:1007–28
- Carlsson H. and E. van Damme. 1993. 'Global games and equilibrium selection'. *Econometrica*, 61:989–1018
- Colman, A. and M. Bacharach. 1997. 'Pay-off dominance and the Stackelberg heuristic'. *Theory and Decision* 43:1–19
- Crawford, V. and H. Haller. 1990. 'Learning how to cooperate: optimal play in repeated coordination games'. *Econometrica*, 58:3571–95
- Gauthier, D. 1975. 'Coordination'. *Dialogue*, 14:195–221
- Gauthier, D. 1986. *Morals by Agreement*, Oxford University Press
- Gilbert, M. 1989. 'Rationality and salience'. *Philosophical Studies*, 57:61–77
- Gilbert, M. 1990. 'Rationality, coordination and convention'. *Synthese*, 84:1–21
- Goyal, S. and M. Janssen. 1996. 'Can we rationally learn to coordinate?' *Theory and Decision*, 40:29–49
- Hollis, M. 1998. *Trust within Reason*. Cambridge University Press
- Jacobsen, H. 1996. 'On the foundations of Nash Equilibrium'. *Economics and Philosophy* 12:67–88
- Janssen, M. 1998a. 'Focal points'. In *The New Palgrave Dictionary of Economics and the Law*, Vol. II, pp. 150–5, P. Newman (ed.). MacMillan
- Janssen, M. 1998b. 'Individualism and equilibrium coordination in games'. In *Crossing Boundaries: Case studies in Economic Methodology*, pp. 1–35. R. Backhouse, D. Hausman, U. Mäki and A. Salanti (eds.). MacMillan
- Janssen, M. 2000. 'Rationalizing focal points'. *Theory and Decision* (forthcoming)
- Kandori, M., G. Malaith and R. Rob. 1993. 'Learning, mutation and long run equilibria in games'. *Econometrica*, 61:29–56
- Kreps, D. 1990. *A Course in Microeconomic Theory*. Harvester Wheatsheaf
- Lewis, D. 1969. *Convention*. Harvard University Press
- Mehta, J., C. Starmer and R. Sugden. 1994a. 'Focal points in pure coordination games: an experimental investigation'. *Theory and Decision*, 36:163–85
- Mehta, J., C. Starmer and R. Sugden. 1994b. 'The nature of salience: an experimental investigation in pure coordination games'. *American Economic Review*, 84:658–73
- Pearce D. 1984. 'Rationalizable strategic behavior and the problem of perfection'. *Econometrica* 52:1029–1050
- Rasmusen, E. 1989. *Games and Information*. Basil Blackwell
- Schelling, T. 1960. *The Strategy of Conflict*. Harvard University Press
- Sugden, R. 1991. 'Rational choice: a survey of contributions from economics and philosophy'. *Economic Journal* 101:751–86
- Sugden, R. 1995. 'Towards a theory of focal points'. *Economic Journal* 105:533–550
- Sugden, R. 2000. 'Team preferences'. *Economics and Philosophy*, 16:175–204
- Young, H. P. 1993. 'The evolution of conventions'. *Econometrica*, 61:57–84