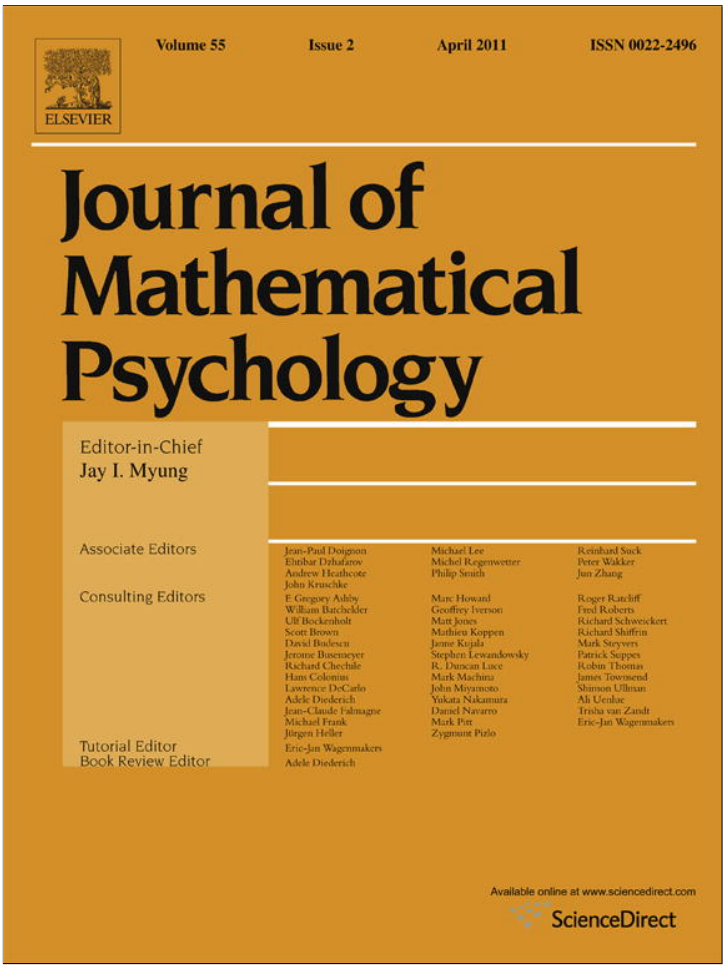

Journal of Mathematical Psychology 55 (2011) 166–175

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

# Mutual support in games: Some properties of Berge equilibria☆

Andrew M. Colman [a], Tom W. Körner [b], Olivier Musy [c], Tarik Tazdaït [d,*]

[a] School of Psychology, University of Leicester, Leicester LE1 7RH, United Kingdom
[b] Department of Pure Mathematics and Mathematical Statistics, Wilberforce Road, Cambridge CB3 0WA, United Kingdom
[c] Université Paris Ouest Nanterre, La Défense, Economix, 200 avenue de la République, 92001 Nanterre Cedex, France
[d] CNRS-EHESS-CIRED, Campus du Jardin Tropical - 45 bis, avenue de la Belle Gabrielle, 94736 Nogent Sur Marne Cedex, France

ARTICLE INFO

*Article history:*
Received 22 December 2009
Received in revised form
16 November 2010
Available online 25 February 2011

*Keywords:*
Altruism
Berge equilibrium
Common interest game
Cooperation
Coordination game
Payoff dominance
Prisoner's dilemma
Social dilemma
Social value orientation
Team reasoning

ABSTRACT

The Berge equilibrium concept formalizes mutual support among players motivated by the altruistic social value orientation in games. We prove some basic results for Berge equilibria and their relations to Nash equilibria, and we provide a straightforward method for finding Berge equilibria in $n$-player games. We explore some specific examples, and we explain how the Berge equilibrium provides a compelling model of cooperation in social dilemmas. We show that the Berge equilibrium also explains coordination in some common interest games and is partially successful in explaining the payoff dominance phenomenon, and we comment that the theory of team reasoning provides alternative solutions to these problems.



## 1. Introduction

This article focuses on the Berge equilibrium concept and explores its relevance to cooperation in social dilemmas and to the related phenomenon of coordination in common interest games. These are among the most familiar problems in social, political, and economic life, and they are poorly understood, although cooperation, including its evolution and maintenance, is arguably "the most important unanswered question in evolutionary biology, and more generally in the social sciences" (May, 2006, p. 109). Berge equilibrium can be viewed as an implication of the altruistic social value orientation of interdependence theory, just as Nash equilibrium is an implication of the individualistic orientation. Berge equilibrium has not previously been explored in the social and behavioral sciences, although it offers potentially useful insights into altruism, cooperation, and coordination.

A fundamental assumption of classical game theory, and of decision theory in general, is that decision makers are invariably motivated to maximize their individual utilities, relative to their knowledge and beliefs at the time of acting. Recent developments in psychological and behavioral game theory have shown that this assumption cannot be true if individual utilities are interpreted narrowly as objective payoff gains or losses, measured in monetary units, for example. Players in strategic games do not invariably try to maximize their individual objective payoffs but sometimes appear to be motivated by *other-regarding utilities* that take account of the payoffs to their co-players. In particular, it is now widely acknowledged that considerations of fairness and reciprocity influence strategy choices, and various implications of this have been explored by Arnsperger and Varoufakis (2003), Bolton and Ockenfels (2000), Fehr and Schmidt (1999), Rabin (1993) and others.

In the Prisoner's Dilemma game, experimental investigations have provided overwhelming evidence that human decision makers do not invariably choose the Nash equilibrium strategies that are mandated by game theory. In the simplest type of Prisoner's Dilemma game, two players each choose between a cooperative strategy, providing a benefit $b$ to the co-player at a cost $c$ to the cooperator, and a defecting strategy that entails no benefit or cost to either player. Provided that $c < b$, this payoff

☆ We wish to thank Denis Bouyssou, Bertrand Crettez, Regis Deloche, Patrice Dumas, Moussa Larbani, Rabia Nessah, Daniel Théry, Yanis Varoufakis, and three anonymous reviewers for helpful comments on previous versions of this article.

* Corresponding author.
*E-mail addresses:* amc@le.ac.uk (A.M. Colman), T.W.Korner@dpmms.cam.ac.uk (T.W. Körner), olivier-musy@wanadoo.fr (O. Musy), tazdait@centre-cired.fr (T. Tazdaït).

0022-2496/$ – see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmp.2011.02.001

function yields a simple (decomposable) version of the familiar Prisoner's Dilemma game. Mutual cooperation results in each player receiving a net payoff of $b - c$, and this is preferable to the zero payoff for mutual defection. But both players nevertheless have a temptation to defect, because a unilateral defector receives the best payoff – the benefit $b$ without any cost – while a unilateral cooperator receives the worst payoff, paying the cost $c$ without any benefit. (This method of specifying a Prisoner's Dilemma game highlights the fact that cooperation can be viewed as mutual altruism, where altruism is defined in the usual way as acting to benefit another individual at some cost to oneself.) Defecting is a *dominant strategy* in the sense that it yields a better payoff irrespective of whether the co-player cooperates or defects, and it follows that this is the only rational way to play an unrepeated or one-shot Prisoner's Dilemma game. What makes the game a genuine dilemma is the fact that each player receives a better individual payoff if both cooperate than if both defect. Despite the dominance of the defecting strategy, people cooperate frequently in experimental Prisoner's Dilemma games and in multiplayer versions of the game (Colman, 2003). In the closely related Centipede game, most players also avoid the (subgame-perfect) Nash equilibrium and behave more cooperatively or altruistically, even when very large financial incentives are at stake (Parco, Rapoport, & Stein, 2002); and in the Ultimatum game, players almost *always* behave more cooperatively or altruistically than is required by the subgame-perfect Nash equilibrium (Camerer & Thaler, 1995).

In all such games, intuition and experimental evidence strongly favor strategy choices that deviate systematically from those mandated by motivations that are purely selfish in terms of objective payoffs. That decision makers are invariably motivated to maximize their individual objective payoffs is a tacit assumption so deeply embedded in the judgment and decision making research tradition that researchers sometimes have difficulty even recognizing that they are making it, but there are circumstances in which even pure altruism seems entirely natural. For a homely example, a doting grandparent playing a board game with a child might be motivated solely to maximize the child's payoffs in the game and may therefore play to lose—technically, playing the game in *misère* mode. It is not difficult to think of interactive decisions in which, on commonsense grounds, we should expect both or all players to choose altruistic strategies. As a benchmark example for this article, consider a jazz-loving man married to a classical music lover. Suppose that each wishes to choose a musical recording as a wedding anniversary present for the other, knowing that they will inevitably spend many hours listening to the music together. Here, we should intuitively expect each spouse to maximize the other's payoff by choosing the other's favorite type of music. In game-theoretic terms, both players would still be maximizing their own individual utilities, as required by expected utility theory, but those utilities would be altruistic rather than selfish.

Although altruism may be relatively uncommon in everyday life, there is extensive empirical evidence that it can be elicited reliably by exposing research participants to certain circumstances, such as the plight of a suffering victim calculated to elicit empathic emotions (see Batson & Shaw, 1991, for a critical review of many of the classic studies). For example, Batson and Ahmad (2001) asked participants to choose second in a one-shot Prisoner Dilemma game, knowing that the co-player had already defected. In a treatment condition in which participants were induced to feel empathy for the co-player who had recently suffered a painful relationship break-up, 45% cooperated, compared to 0% in a control condition. Among the other factors that are known to be associated with altruism is reciprocity: in many types of interaction, altruistic behavior is essentially linked to an expectation of reciprocal altruism from the recipient (Brosnan & de Waal, 2002; Trivers, 2005).

Altruistic behavior is well documented in nonhuman species. A familiar example observed in many species of birds is the distinctive alarm call that they emit when they spot predators, such as hawks (Maynard Smith, 1965). Alarm calls alert other members of the flock, enabling them to take appropriate evasive action, but such behavior provides no benefit to the alarm-caller itself; on the contrary, there is persuasive (though indirect) evidence that, by attracting the predator's attention, the alarm-caller actually *reduces* its own chances of survival (Marler, 1955, 1959). It is now widely acknowledged that alarm calling provides a true example of altruistic behavior in nature (Wilson & Evans, 2008).

*Interdependence theory* is based on the undeniable premise that the utilities determining players' strategy choices in games do not invariably correspond to their objective payoffs (for reviews, see Rusbult & Van Lange, 2003; Van Lange, 2000). The motivations of players in two-player games are described in terms of *social value orientations* defined by *payoff transformations*. According to this approach, players' other-regarding utilities $U'$ are defined as functions of their own and their co-players' objective payoffs $U$. If $u_i$ and $u_j$ are the objective payoffs to Players $i$ and $j$ in a two-player game, and $s_i$ and $s_j$ are strategies chosen by Players $i$ and $j$, then Player $i$ is assumed to maximize a real-valued utility function $U'_i(s_i, s_j) = f_i(u_i, u_j)$, and Player $i$'s social value orientation is a property of the particular function $f_i$. The *individualistic* orientation is represented by $f_i = u_i$, the *altruistic* orientation by $f_i = u_j$, the *cooperative* orientation by $f_i = u_i + u_j$, the *competitive* orientation by $f_i = u_i - u_j$, and the *equality-seeking* orientation by $f_i = \min\{u_i - u_j, u_j - u_i\}$. The assumption is that players are invariably motivated to maximize their expected utilities, in any situations in which they find themselves, but that that these expected utilities are not necessarily individualistic—they may be altruistic, cooperative, competitive, or equality-seeking, depending on the psychological characteristics of the decision maker and the particular circumstances of the social interaction. We shall show that Berge equilibria arise in circumstances in which utility-maximizing players are motivated by the altruistic social value orientation.

The individualistic social value orientation is the motivation tacitly assumed by decision theory and game theory, and the voluminous literature of orthodox game theory may be viewed as an exploration of its theoretical implications. The remainder of this article is devoted to a preliminary exploration of Berge equilibrium as a theoretical implication of the altruistic social value orientation. In Section 2, we formalize Berge equilibrium, examine some examples, derive some basic theoretical results, and propose a new model of cooperation in social dilemmas in terms of Berge equilibrium. In Section 3, we show how the Berge equilibrium solves coordination problems in some common interest games and partially explains the payoff dominance problem, and we argue that the cooperative social value orientation, in conjunction with the theory of team reasoning, offers a complete solution to these problems. In Section 4 we briefly review the sparse literature on Berge equilibrium and social value orientations. In Section 5 we draw the threads together and summarize our conclusions.

## 2. Altruism and the Berge equilibrium

The Berge equilibrium concept was introduced intuitively by Berge (1957, p. 20) and formalized by Zhukovskii (1985) in the context of differential games. Early literature, almost exclusively in Russian, focused on differential games. Berge equilibrium was not examined in conventional strategic games until recently (Abalo & Kostreva, 2004, 2005), and emerging recognition of other-regarding utilities in psychological and behavioral game theory (Bolton & Ockenfels, 2000; Fehr & Schmidt, 1999) has imbued it with contemporary relevance.

We consider a game[1]

$$G := \langle N, (S_i)_{i\in N}, (U_i)_{i\in N}\rangle,$$

where $N = \{1, 2, \ldots, n\}$ is the set of players, $S_i$ denotes the strategy set of Player $i$, and $U_i$ is the utility function of Player $i$. We assume that $n \geq 2$ and that each $S_i$ is finite and contains at least two strategies. We write $S = \prod_{i\in N} S_i$, and we interpret $U_i$ as a function $U_i : S \to \mathbb{R}$ representing the Player $i$'s objective payoffs. We call $\mathbf{s} = (s_1, s_2, \ldots, s_n) \in S$ a *strategy profile*, but we shall also be interested in the *incomplete strategy profile*

$$\mathbf{s}_{-i} = (s_1, s_2, \ldots, s_{i-1}, s_{i+1}, \ldots, s_n) \in \prod_{j\neq i} S_j.$$

We use the natural notation

$$U_i(s_i, \mathbf{s}_{-i}) = U(s_1, s_2, \ldots, s_{i-1}, s_i, s_{i+1}, \ldots, s_n).$$

We call a strategy profile $\mathbf{s}^* \in S$ a *Berge equilibrium* if, writing $\mathbf{s}^* = (s_1^*, s_2^*, \ldots, s_n^*)$, we have

$$U_i(s_i^*, \mathbf{s}_{-i}) \leq U_i(\mathbf{s}^*) \tag{1}$$

for all $\mathbf{s}_{-i} \in S_{-i}$ and for all $i \in N$. In other words, if the players have chosen a strategy profile that forms a Berge equilibrium, and $i$ sticks to the chosen strategy but some of the other players change their strategies, then $i$'s payoff will not increase.

For Nash equilibrium (Nash, 1950, 1951), we have

$$U_i(s_i, \mathbf{s}_{-i}^*) \leq U_i(\mathbf{s}^*)$$

for all $s_i \in S_i$ and for all $i \in N$. A Nash strategy is a *best reply* to co-players who also play Nash strategies, yielding the best payoff to a player who chooses it, given the strategies chosen by the co-player(s), whereas a Berge strategy yields the best payoffs to co-players who also play Berge strategies. The key difference is that, in Nash equilibrium, an individual player's deviation from the equilibrium can reduce that player's own payoff whereas, in Berge equilibrium, a deviation by one or more co-players can reduce the payoff to a player who does not deviate.

Zhukovskii and Chikrii (1994) specify a further condition. Let us write

$$\alpha_i = \max_{s_i\in S_i} \min_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i, \mathbf{s}_{-i}),$$

so that $\alpha_i$ is $i$'s *maximin security level*, that is, the maximum payoff that $i$ can guarantee to obtain regardless of the choices of the other players. Zhukovskii and Chikrii stipulate that

$$\alpha_i \leq U_i(\mathbf{s}^*) \tag{2}$$

for all $i \in N$. If a Berge equilibrium strategy profile $\mathbf{s}^*$ also obeys (2), then we shall call $\mathbf{s}^*$ a *Berge–Vaisman equilibrium* (for reasons to be explained in Section 4).

In interdependence theory, the altruistic social value orientation, in which Player $i$ is assumed to maximize the utility function $U_i'(s_i, s_j) = f_i(u_i, u_j) = u_j$, has been defined only for two-player games. Berge equilibrium may reasonably be interpreted as an implication of the altruistic orientation, generalized to $n$-player games with otherwise standard game-theoretic assumptions, because in both cases every player is motivated to maximize the payoff(s) to the other player(s) and every player chooses strategies accordingly.

We illustrate these ideas by finding a simple sufficient condition for a Berge equilibrium to be a Berge–Vaisman equilibrium. We need the following version of a standard lemma.

[1] We exclude from consideration mixed extensions in which players can choose probability distributions over their pure strategy sets. This article focuses on elementary games in which players can choose pure strategies only.

**Lemma 1.** *With the notation that we have introduced,*

$$\max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}) \leq \min_{s_i\in S_i} \max_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i, \mathbf{s}_{-i})$$

*for all $i \in N$.*

**Proof.** Fix $i \in N$. We have

$$U_i(s_i, \mathbf{s}_{-i}) \leq \max_{\mathbf{t}_{-i}\in S_{-i}} U_i(s_i, \mathbf{t}_{-i})$$

for all $s_i \in S_i$. It follows that

$$\min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}) \leq \min_{s_i\in S_i} \max_{\mathbf{t}_{-i}\in S_{-i}} U_i(s_i, \mathbf{t}_{-i}) = \min_{t_i\in S_i} \max_{\mathbf{t}_{-i}\in S_{-i}} U_i(t_i, \mathbf{t}_{-i})$$

and

$$\max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}) \leq \min_{t_i\in S_i} \max_{\mathbf{t}_{-i}\in S_{-i}} U_i(t_i, \mathbf{t}_{-i}).$$

Substituting $s_i$ for $t_i$ and $\mathbf{s}_i$ for $\mathbf{t}_i$, we obtain

$$\max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}) \leq \min_{s_i\in S_i} \max_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i, \mathbf{s}_{-i}). \quad \square$$

Having examined $\alpha$ as defined in (2), it is natural also to examine

$$\beta_i = \max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}).$$

The following theorem tells us that, if $\alpha_i \leq \beta_i$ for all $i$, then a Berge equilibrium is a Berge–Vaisman equilibrium.

**Theorem 1.** *Suppose that the game discussed above has*

$$\max_{s_i\in S_i} \min_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i, \mathbf{s}_{-i}) \leq \max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i})$$

*for all $i \in N$. Then any Berge equilibrium is a Berge–Vaisman equilibrium.*

**Proof.** Suppose that $\mathbf{s}^*$ is a Berge equilibrium. By definition,

$$U_i(s_i^*, \mathbf{s}_{-i}) \leq U_i(\mathbf{s}^*)$$

for all $\mathbf{s}_{-i} \in S_{-i}$ and all $i \in N$. It follows that

$$\max_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i^*, \mathbf{s}_{-i}) \leq U_i(\mathbf{s}^*),$$

and consequently

$$\min_{s_i\in S_i} \max_{\mathbf{s}_{-i}\in S_{-i}} U_i(s_i, \mathbf{s}_{-i}) \leq U_i(\mathbf{s}^*).$$

Applying Lemma 1, we obtain

$$\max_{\mathbf{s}_{-i}\in S_{-i}} \min_{s_i\in S_i} U_i(s_i, \mathbf{s}_{-i}) \leq U_i(\mathbf{s}^*),$$

or, in more abbreviated notation, $\beta_i \leq U_i(\mathbf{s}^*)$. If $\alpha_i \leq \beta_i$ for all $i \in N$, then $\alpha_i \leq U_i(\mathbf{s}^*)$ for all $i$, and therefore $\mathbf{s}^*$ is a Berge–Vaisman equilibrium. $\square$

If Player $i$ adopts a Berge strategy, then why should the other altruistically motivated players follow suit by playing their own Berge strategies? Consider a player $j$ who plays a non-Berge strategy $s_j \neq s_j^*$. According to the defining condition (1), we have:

$$U_i(s_i^*, \mathbf{s}_{-i-j}^*, s_j^*) \geq U_i(s_i^*, \mathbf{s}_{-i-j}^*, s_j)$$

for all $i \in N$ and $s_j \neq s_j^*$. By playing a Berge strategy, $j$ maximizes $i$'s utility and, because this applies to every $i \in N, j$ maximizes the utility to all the other players in the game as well. Furthermore, by playing Berge strategies, the other players also maximize $j$'s utility. In other words, in a Berge equilibrium, every player $i$ maximizes the utilities of the co-players $N \setminus \{i\}$, and $i$'s utility is simultaneously maximized by those co-players. In a Berge–Vaisman equilibrium, because of Condition

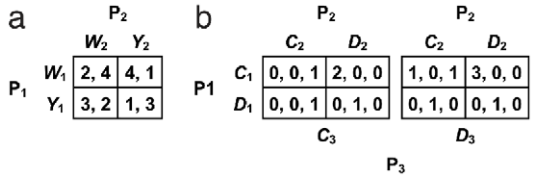


**Fig. 1.** Two games without Berge equilibria: (a) Gibbard's two-player game; (b) an arbitrary three-player game.

2, it is also the case that a selfishly motivated player could not obtain a better payoff by defecting unilaterally. Given standard game-theoretic knowledge and rationality assumptions, this stronger Berge–Vaisman equilibrium can therefore arise even if it is common knowledge in a game that every player but one is altruistically motivated and one is selfishly motivated. But reciprocity in the absence of altruism cannot explain the Berge–Vaisman equilibrium on its own, because if *all* players are selfishly motivated, then they will be motivated to select Nash rather than Berge–Vaisman equilibria.

### 2.1. Examples of Berge equilibria

Not all games have Berge equilibria, as shown by the following two-player and three-player games. Fig. 1(a) shows Gibbard's (1974) game, a 2 × 2 game lacking (in pure strategies) both Nash equilibria and Berge equilibria. Gibbard's interpretation is of two players living together and choosing either white ($W$) or yellow ($Y$) as a color to decorate their bedrooms. Player 1 is a nonconformist who prefers to have a different color from Player 2, whereas Player 2 is a conformist who prefers to have the same color as Player 1; and secondary to these preferences, each player prefers white to yellow. Representing the players' preferences from best to worst by the ordinal numbers 4, 3, 2, and 1, the payoffs are as shown in Fig. 1(a). The lack of any Berge equilibrium is confirmed by the following strict inequalities, covering all four possible outcomes, each violating Condition (1) in the definition of the Berge equilibrium above:

$U_1(W_1, W_2) < U_1(W_1, Y_2)$;
$U_2(W_1, Y_2) < U_2(Y_1, Y_2)$;
$U_2(Y_1, W_2) < U_2(W_1, W_2)$;
$U_1(Y_1, Y_2) < U_1(Y_1, W_2)$.

Fig. 1(b) shows a 2 × 2 × 2 game with Nash equilibria at $(C_1, C_2, C_3)$, $(C_1, D_2, C_3)$, $(C_1, C_2, D_3)$, and $(C_1, D_2, D_3)$. In spite of this plethora of Nash equilibria, the game has no Berge equilibrium, because each of the following strict inequalities violates Condition (1):

$U_1(C_1, C_2, C_3) < U_1(C_1, C_2, D_3)$;
$U_1(C_1, C_2, D_3) < U_1(C_1, D_2, C_3)$;
$U_1(C_1, D_2, C_3) < U_1(C_1, D_2, D_3)$;
$U_2(C_1, D_2, D_3) < U_2(D_1, D_2, C_3)$;
$U_2(D_1, C_2, C_3) < U_2(D_1, C_2, D_3)$;
$U_3(D_1, C_2, D_3) < U_3(C_1, C_2, D_3)$;
$U_3(D_1, D_2, C_3) < U_3(C_1, C_2, C_3)$;
$U_3(D_1, D_2, D_3) < U_3(C_1, C_2, D_3)$.

The first inequality shows that the outcome $(C_1, C_2, C_3)$ cannot be a Berge equilibrium, because Player 1 receives a better payoff in the outcome $(C_1, C_2, D_3)$, where Player 3 has switched to $D$, violating Condition (1), and so on for each of the eight possible outcomes of the game in order, forcing the conclusion that the game has no Berge equilibrium.

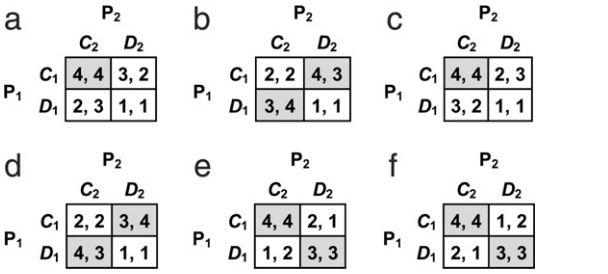


**Fig. 2.** Ordinally distinct symmetric 2 × 2 games in which Nash and Berge equilibria coincide. The Nash–Berge equilibria are shaded.

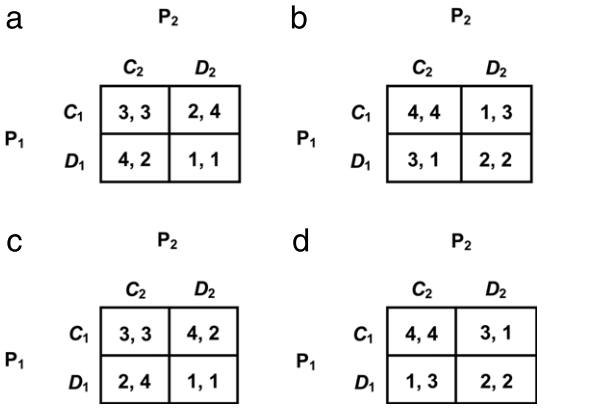


**Fig. 3.** Ordinally distinct symmetric 2 × 2 games with multiple Nash equilibria and unique Berge equilibria (a and b) or unique Nash equilibria and multiple Berge equilibria (c and d).

Restricting attention to symmetric 2 × 2 games in which both players have strict preferences among the four outcomes, and representing their payoffs once again by the ordinal numbers 4, 3, 2, and 1, there are exactly 12 such *ordinally distinct* symmetric 2 × 2 games to consider (Rapoport, Guyer, & Gordon, 1976). It turns out that each of these games has at least one Nash equilibrium and at least one Berge equilibrium, as detailed below.

In six cases, Nash and Berge equilibria coincide, as shown in Fig. 2, where the Nash–Berge equilibria are shaded. Some of these games are sufficiently well known to have familiar names: Rapoport (1967) originally named Fig. 2(b) Hero, and Fig. 2(d) Leader; and Sen (1969) named Fig. 2(f) the Assurance game.

Four of the games have multiple Nash equilibria and unique Berge equilibria, or vice versa, as shown in Fig. 3. Fig. 3(a) is sometimes called the Chicken game, after an interpretation by the philosopher Bertrand Russell (1959, p. 30) of a dangerous game in which two motorists speed toward each other on a head-on collision course, and if either player swerves out of the way, then that player is shown to be "chicken". It is sometimes called the Hawk-Dove game, after a biological interpretation by Maynard Smith and Price (1973) of conventional and escalated fighting in animals, or the Snowdrift game, after a scenario suggested by Sugden (1986, p. 128) in which two motorists on a lonely road, both equipped with shovels, are stuck in a snowdrift. It has Nash equilibria at $(D_1, C_2)$ and $(C_1, D_2)$ and a unique Berge equilibrium at $(C_1, C_2)$. Fig. 3(b) is the Stag Hunt game, named after a passage in Rousseau (1755, Part II, paragraph 9), in which two hunters have to cooperate to catch a stag, but each is tempted to chase after the lesser prize of a hare, which can be caught without the other's cooperation, and it was introduced into the literature of game theory by Lewis (1969, p. 7); it has Nash equilibria at $(C_1, C_2)$ and $(D_1, D_2)$ and a unique Berge equilibrium at $(C_1, C_2)$. Fig. 3(c) shows a game with a unique Nash equilibrium at $(C_1, C_2)$ and multiple Berge equilibria at $(C_1, D_2)$ and $(D_1, C_2)$. The game in Fig. 3(d) has a

a

| P1 \ P2 | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 3, 3 | 4, 1 |
| $D_1$ | 1, 4 | 2, 2 |

b

| P1 \ P2 | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 3, 3 | 1, 4 |
| $D_1$ | 4, 1 | 2, 2 |

**Fig. 4.** Ordinally distinct symmetric $2 \times 2$ games with unique Nash equilibria and non-coinciding unique Berge equilibria.

unique Nash equilibrium at $(C_1, C_2)$ and multiple Berge equilibria at $(C_1, C_2)$ and $(D_1, D_2)$.

The two remaining cases, shown in Fig. 4, are games with unique Nash equilibria and unique but distinct Berge equilibria. Game 4(a) has a unique Nash equilibrium at $(C_1, C_2)$ and a unique Berge equilibrium at $(D_1, D_2)$. Game 4(b) is the Prisoner's Dilemma game, first named by Tucker (unpublished notes), with a unique Nash equilibrium at $(D_1, D_2)$ and a unique Berge equilibrium at $(C_1, C_2)$.

It is clear that Berge equilibria are sometimes more numerous than Nash equilibria, sometimes less numerous, and sometimes equally numerous, although even in the last case Berge and Nash equilibria do not necessarily coincide. Our exhaustive survey of ordinally distinct, symmetric $2 \times 2$ games gives an indication of how common Berge equilibria are in these small games. In Section 2.3, we shall show that Berge equilibria are ubiquitous in a class of $n$-player games frequently encountered in social, economic, and political interactions in everyday life.

### 2.2. Relations between Berge and Nash equilibria

We now examine the relation between Berge equilibria and Nash equilibria. Recall that, with our usual notation, $\mathbf{s}^*$ is a Nash equilibrium if

$$U_i\left(s_i, \mathbf{s}^*_{-i}\right) \leq U_i\left(\mathbf{s}^*\right) \tag{3}$$

for all $s_i \in S_i$ and all $i \in N$. In other words, if the players have chosen a Nash equilibrium strategy profile, and if $i$ changes to a different strategy and the other players do not, then $i$'s payoff will not increase.

In the particular case of two-player games, there is a high degree of symmetry between the Nash and Berge conditions. Intuitively, players in a two-player game may be thought to interchange their payoff functions, in some sense, when they switch between Berge and Nash equilibria. We clarify and formalize this idea in the following lemma.

**Lemma 2.** *The strategy profile* $\mathbf{s}^*$ *in the two-player game*

$$G := \langle \{1, 2\}, (S_1, S_2), (U_1, U_2) \rangle$$

*is a Berge equilibrium if and only if it is a Nash equilibrium in the game*

$$G := \langle \{1, 2\}, (S_1, S_2), (V_1, V_2) \rangle$$

*with* $V_1 = U_2$ *and* $V_2 = U_1$.

**Proof.** It suffices to observe that the statement

$$U_1\left(s_1, s_2^*\right) \leq U_1\left(s_1^*, s_2^*\right) \quad \text{and} \quad U_2\left(s_1^*, s_2\right) \leq U_2\left(s_1^*, s_2^*\right)$$

for all $s_1 \in S_1, s_2 \in S_2$, and the statement

$$V_2\left(s_1, s_2^*\right) \leq V_2\left(s_1^*, s_2^*\right) \quad \text{and} \quad V_1\left(s_1^*, s_2\right) \leq V_1\left(s_1^*, s_2^*\right)$$

for all $s_1 \in S_1, s_2 \in S_2$, are equivalent. □

We now extend part of this result to $n$-player games, but first we need to introduce some additional notation. If $\sigma : N \to N$ is a permutation, then we say that $\sigma$ is a *derangement* if $\sigma(i) \neq i$ for all $i \in N$. If $G$ is our general game

$$G := \left\langle N, (S_i)_{i \in N}, (U_i)_{i \in N} \right\rangle,$$

and $\sigma$ is a derangement, we can define an associated game

$$G_\sigma := \left\langle N, (S_i)_{i \in N}, (V_i)_{i \in N} \right\rangle$$

with $V_i = U_{\sigma(i)}$. (Here, once again, players interchange payoff functions.)

**Theorem 2.** *If* $\mathbf{s}^*$ *is a Berge equilibrium in the game*

$$G := \left\langle N, (S_i)_{i \in N}, (U_i)_{i \in N} \right\rangle,$$

*then it is a Nash equilibrium in every game* $G_\sigma$ *with derangement* $\sigma$.

**Proof.** Suppose that $\mathbf{s}^*$ is a Berge equilibrium in $G$. Fix $i \in N$ and let $\sigma$ be a derangement. If we can show that

$$U_{\sigma(i)}\left(s_i, \mathbf{s}^*_{-i}\right) \leq U_{\sigma(i)}\left(\mathbf{s}^*\right)$$

for all $s_i \in S_i$ and all derangements $\sigma$, then the required result will follow. By symmetry, we need only consider the case $i = 1, \sigma(i) = 2$.

By the definition of Berge equilibrium,

$$U_2\left(s_2^*, \mathbf{s}_{-2}\right) \leq U_2\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-2} \in S_{-2}$. Thus

$$U_2\left(s_1, s_2^*, s_3, \ldots, s_n\right) \leq U_2\left(s_1^*, s_2^*, s_3^*, \ldots, s_n^*\right)$$

for all $s_k \in S_k$ and all $k \in N, k \neq 2$. In particular,

$$U_2\left(s_1, s_2^*, s_3^*, \ldots, s_n^*\right) \leq U_2\left(s_1^*, s_2^*, s_3^*, \ldots, s_n^*\right)$$

for all $s_1 \in S_1$, and the required result follows. □

If we write $E_\sigma$ for the set of Nash equilibria of $G_\sigma$ and $B$ for the set of Berge equilibria of $G$, then we have proved that

$$B \subseteq \bigcap E_\sigma,$$

where $\sigma$ ranges over all derangements of $N$. The following corollary follows immediately.

**Corollary 1.** *If any of the games* $G_\sigma$ *with derangement* $\sigma$ *has no Nash equilibrium, then the corresponding game* $G$ *has no Berge equilibrium and* $B = \emptyset$.

We now provide a link between Berge equilibria in $n$-player games and Nash equilibria in some associated two-player games. Consider our general game

$$G := \left\langle N, (S_i)_{i \in N}, (U_i)_{i \in N} \right\rangle.$$

For every $j \in N$, we define an associated two-player game

$$G_j := \langle \{A, C\}, (T_A(j), T_C(j)), (V_A(j), V_C(j)) \rangle,$$

with

$$T_A(j) = S_j, T_C(j) = S_{-j},$$

and

$$V_A(j)\left(s_j, \mathbf{s}_{-j}\right) = \sum_{i \neq j} U_i\left(s_j, \mathbf{s}_{-j}\right), V_C(j)\left(s_j, \mathbf{s}_{-j}\right) = U_j\left(s_j, \mathbf{s}_{-j}\right).$$

We write $E_j$ for the set of Nash equilibria in the game $G_j$.

**Theorem 3.** *With the notation just introduced,*

$$B = \bigcap_{j \in N} E_j.$$

**Proof.** We begin by showing that $B \subseteq E_j$ for every $j \in N$. By symmetry, it suffices to show that $B \subseteq E_1$. We write $T_A = T_A(1), T_C = T_C(1), V_A = V_A(1)$, and $V_C = V_C(1)$.

Suppose $\mathbf{s}^* \in B$. By definition of Berge equilibrium, we have

$$U_i\left(s_i^*, \mathbf{s}_{-i}\right) \leq U_i\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-i} \in S_{-i}$ and all $i \in N$. Rewriting, we have

$$U_i\left(s_1, s_2, \ldots, s_{i-1}, s_i^*, s_{i+1}, \ldots, s_n\right) \leq U_i\left(\mathbf{s}^*\right)$$

for all $s_k \in S_k$ $(k \neq i)$. In particular,

$$U_i\left(s_1, s_2^*, \ldots, s_{i-1}^*, s_i^*, s_{i+1}^*, \ldots, s_n^*\right) \leq U_i\left(\mathbf{s}^*\right)$$

for all $s_1 \in S_1$ and $i \neq 1$. Summing over $i \neq 1$, we get

$$\sum_{i\neq 1} U_i\left(s_1, s_2^*, \ldots, s_{i-1}^*, s_i^*, s_{i+1}^*, \ldots, s_n^*\right) \leq \sum_{i\neq 1} U_i\left(\mathbf{s}^*\right),$$

and it follows that

$$V_A\left(s_1, \mathbf{s}_{-1}^*\right) \leq V_A\left(\mathbf{s}^*\right)$$

for all $s_1 \in S_1$.

On the other hand, because

$$U_i\left(s_i^*, \mathbf{s}_{-i}\right) \leq U_i\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-i} \in S_{-i}$ and all $i \in N$, we have, in particular,

$$U_1\left(s_1^*, \mathbf{s}_{-1}\right) \leq U_1\left(\mathbf{s}^*\right).$$

Therefore

$$V_C\left(s_1^*, \mathbf{s}_{-1}\right) \leq V_C\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-1} \in S_{-1}$. This shows that $\mathbf{s}^*$ is a Nash equilibrium of $G_1$, as required.

We have shown that $B \subseteq \bigcap_{j\in N} E_j$. Now suppose that $\mathbf{s}^* \in \bigcap_{j\in N} E_j$. We wish to show that $\mathbf{s}^* \in B$, that is

$$U_i\left(s_i^*, \mathbf{s}_{-i}\right) \leq U_i\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-i} \in S_{-i}$ and all $i \in N$. By symmetry, it suffices to show that

$$U_1\left(s_1^*, \mathbf{s}_{-i}\right) \leq U_1\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-1} \in S_{-1}$. But, taking $V_C = V_C(1)$, this is precisely the statement

$$V_C\left(s_1^*, \mathbf{s}_{-1}\right) \leq V_C\left(\mathbf{s}^*\right)$$

for all $\mathbf{s}_{-1} \in S_{-1}$, guaranteed by the fact that $\mathbf{s}^* \in E_1$. It follows that

$$B = \bigcap_{j\in N} E_j. \quad \square$$

**Remark.** Theorem 3 establishes that the Berge equilibria of an $n$-player game $G$ can be determined by finding the common Nash equilibria in the associated two-player games $G_\sigma$. Well known procedures are available for finding Nash equilibria in two-player games; therefore Theorem 3 suggests a straightforward procedure for finding Berge equilibria in $n$-player games. Corollary 1 shows that if any of the two-player games $G_i$ has no pure-strategy Nash equilibrium, then the $n$-player game $G$ has no Berge equilibrium.

### 2.3. Interpreting cooperation

The Prisoner's Dilemma game shown in Fig. 4(b) is widely interpreted as a game-theoretic model of cooperation, with the $C$ strategies representing cooperation and the $D$ strategies defection or non-cooperation. It is generally agreed that $(D_1, D_2)$ is the uniquely rational strategy profile, but a large volume of experimental evidence reveals that cooperation is commonplace in this game. In most published experiments, approximately half the observed strategy choices are cooperative, even in experiments with unrepeated or one-shot Prisoner's Dilemmas (Colman, 1995, chap. 7, 2003; Sally, 1995).

**Table 1**
Payoff function for a three-player Prisoner's Dilemma.

| Number choosing $C$ | Number choosing $D$ | Payoff to each $C$-chooser | Payoff to each $D$-chooser |
|---|---|---|---|
| 3 | 0 | 3 | – |
| 2 | 1 | 2 | 4 |
| 1 | 2 | 1 | 3 |
| 0 | 3 | – | 2 |

Whether or not $C$ is interpreted as cooperative or altruistic, Berge equilibrium offers a theoretical model of the empirical findings. If players are altruistically motivated to be mutually supportive, rather than to pursue their individual self-interests in terms of objective payoffs, and if they expect their co-players to be motivated similarly to themselves, then it makes sense for them to choose Berge strategies as a means of achieving that goal. In the Prisoner's Dilemma game, if both players simultaneously choose Berge strategies, then each receives a better payoff than if both choose dominant Nash strategies. With the payoffs shown in Fig. 4(b), for example, each player receives 3 units in Berge equilibrium, but only 2 units in Nash equilibrium. Altruistic mutual support pays better than selfish individualism, provided that both players adopt it.

This property of the Prisoner's Dilemma game generalizes to $n$-player social dilemmas, formalized independently by Hamburger (1973) and Schelling (1973). The payoff functions of a three-player Prisoner's Dilemma are shown in Table 1. A generalized $n$-Player Prisoner's Dilemma $(n \geq 2)$, including the three-player game shown in Table 1, is defined by the following three properties:

(1) Every player's strategy set comprises two strategies: $C$ (cooperate) and $D$ (defect).
(2) The $D$ strategy is strictly dominant for every player—every player obtains a higher payoff by choosing $D$ than $C$, irrespective of the number of co-players choosing $C$.
(3) The dominant $D$ strategies intersect in a Pareto-inefficient Nash equilibrium, the outcome being better for every player if all players choose $C$ than if all choose $D$.

The two-player Prisoner's Dilemma game meets this definition and is therefore a special case of the more general $n$-Player Prisoner's Dilemma. In two-player and $n$-player social dilemmas, joint defection is the unique Nash equilibrium, and joint cooperation the unique Berge equilibrium. In the game shown in Table 1, if all players choose their dominant $D$ strategies, then the payoff to each player is 2 units in Nash equilibrium; but if all players choose their $C$ strategies, then the payoff to each is 3 units in Berge equilibrium. Social dilemmas with the strategic structure of the $n$-Player Prisoner's Dilemma game are ubiquitous in everyday social, political, and economic life, from conservation of scarce resources to wage inflation and global warming. Situations in which individuals are tempted to act selfishly, but everyone ends up worse off if everyone acts selfishly than if everyone cooperates, are remarkably common.

A class of *decomposable* $n$-Player Prisoner's Dilemma games can be characterized as follows. Let each of the $n$ players receive an amount $c$ for choosing $C$ and an amount $d$ for choosing $D$; in addition, let each player lose an amount $e$ for every player in the game who chooses $D$. In the game shown in Table 1, $c = 3$, $d = 5$, and $e = 1$. In the two-player version shown in Fig. 4(b), $c = 3$, $d = 6$, and $e = 2$. We can express the payoff to a $D$-chooser when all co-players choose $C$ as $d - e$, the payoff to a $D$-chooser when all co-players choose $D$ as $d - ne$, the payoff to a $C$-chooser when all co-players choose $C$ as $c$, and the payoff to a $C$-chooser when all co-players choose $D$ as $c - (n-1)e$. The defining properties of the generalized decomposable $n$-Player Prisoner's Dilemma game are the inequalities

$$d - e > c > d - ne > c - (n-1)e.$$

a

| $P_1$ \ $P_2$ | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 9, 9 | 0, 8 |
| $D_1$ | 8, 0 | 7, 7 |

b

| $P_1$ \ $P_2$ | $C_2$ | $D_2$ |
|---|---|---|
| $C_1$ | 18, 18 | 8, 8 |
| $D_1$ | 8, 8 | 14, 14 |

c

| $P_1$ \ $P_2$ | $H_2$ | $L_2$ |
|---|---|---|
| $H_1$ | 2, 2 | 0, 0 |
| $L_1$ | 0, 0 | 1, 1 |

d

| $P_1$ \ $P_2$ | $H_2$ | $L_2$ |
|---|---|---|
| $H_1$ | 4, 4 | 0, 0 |
| $L_1$ | 0, 0 | 2, 2 |

**Fig. 5.** (a) Aumann's Stag Hunt game; (b) Aumann's Stag Hunt game after cooperative payoff transformation; (c) Hi-Lo matching game; (d) Hi-Lo matching game after cooperative payoff transformation.

Simplifying, we find that $n > 1$, which means simply that the number of players must be two or more, and combining the inequalities, we arrive at

$$1 < \frac{d-c}{e} < n,$$

which is a generating formula for a decomposable $n$-Player Prisoner's Dilemma game of any size. Any such game has a unique Nash equilibrium representing joint defection and a unique Berge equilibrium representing joint cooperation. That joint cooperation is a Berge equilibrium follows from the fact that $c > c - ke$ for all $k$ $(0 < k < n)$, satisfying the definition (1) of a Berge equilibrium provided earlier.

A large body of empirical evidence reveals that cooperation is very common, even in unrepeated $n$-Player Prisoner's Dilemma games (Camerer & Fehr, 2006, Colman, 1995, chap. 9, Colman, 2003, Sally, 1995, Suleiman, Budescu, Fischer, & Messick, 2004), and these findings are difficult to understand within the conceptual framework of game theory. In social dilemmas of all sizes, Berge equilibrium offers a model of cooperation that has not been explored by previous researchers. Players motivated by the altruistic social value orientation may select Berge strategies in the expectation that other players will act similarly, knowing that if these expectations are fulfilled, then the mutually supportive players will end up better off, even in terms of objective payoffs, than if they had followed the game-theoretic logic of strategic dominance and Nash equilibrium.

## 3. Coordination in common interest games

One of the most serious limitations of game theory, and a major reason for its notorious indeterminacy, is its frequent failure to provide a rational justification for the selection of *payoff-dominant* Nash equilibria. Consider the game shown in Fig. 5(a), a version of the Stag Hunt game discussed at length by Harsanyi and Selten (1988, pp. 355–363). Both $(C_1, C_2)$ and $(D_1, D_2)$ are Nash equilibria, and both players prefer $(C_1, C_2)$, yielding payoffs of (9, 9), to $(D_1, D_2)$, yielding payoffs of (7, 7). The Nash equilibrium $(C_1, C_2)$ payoff-dominates $(D_1, D_2)$ because it yields a better payoff to both players than does $(D_1, D_2)$ or, to put it differently, because $(C_1, C_2)$ Pareto-dominates $(D_1, D_2)$. Nevertheless, the indeterminacy of game theory is revealed in this game, because game theory provides neither player with any reason to choose the $C$ strategy, in spite of its intuitive appeal. Player $i$ has no reason to choose $C_i$ in the absence of a reason to expect Player $j$ to choose $C_j$.

Formally, if $e$ and $f$ are two Nash equilibria in any game, then $e$ strictly payoff-dominates $f$ if

$$U_i(e) > U_i(f)$$

for all $i \in N$. Any game with multiple Nash equilibria, including one that payoff-dominates all others, is called a *common interest game* (Aumann & Sorin, 1989). The *payoff dominance principle* is the assumption that if an equilibrium $e$ payoff-dominates all others in a common interest game, then rational players will choose and play the component strategies of $e$. Harsanyi and Selten (1988) proposed this principle (together with a secondary *risk dominance* principle) as an axiom of rationality in their general theory of equilibrium selection in games, though only provisionally and reluctantly (see their comments on pp. 362–363), acknowledging that it provides no *explanation* for the powerful intuition that payoff-dominant equilibria should be selected by rational players.

Fig. 5(c) exposes payoff dominance in its simplest and most transparent form. In this Hi-Lo matching game, there are Nash equilibria at $(H_1, H_2)$ and $(L_1, L_2)$, and both players prefer $(H_1, H_2)$. However, Player 1 has no reason to choose $H_1$ in the absence of a reason to expect Player 2 to choose $H_2$, because if Player 2 were to choose $L_2$, for whatever reason, then Player 1's best reply would be $L_1$. But Player 1 has no reason to expect Player 2 to choose $H_2$, because the game is symmetric and Player 2 faces the identical dilemma, hence any attempt to justify such expectations leads to an infinite and inconclusive regress. Orthodox game theory provides no reason for either player to choose the cooperative $H$ strategies.[2]

### 3.1. Berge equilibrium, coordination, and payoff dominance

Does the Berge equilibrium offer a solution to the payoff dominance problem? Surprisingly, although it provides a solution to the Stag Hunt game shown in Fig. 5(a), and to other similar games, it does not solve the Hi-Lo game shown in Fig. 5(c). The Stag Hunt game has a unique Berge equilibrium at $(C_1, C_2)$, but the Hi-Lo game has Berge equilibria coinciding with the Nash equilibria at both $(H_1, H_2)$ and $(L_1, L_2)$, providing no remedy for game-theoretic indeterminacy and no insight into the intuitive appeal of the $H$ strategies in this game.

Attempts to rationalize the payoff-dominant equilibrium $(H_1, H_2)$ in Fig. 5(c) have all resorted to essential changes in the specification of the game, introducing either repetitions (e.g. Aumann & Sorin, 1989), a "cheap talk" stage during which players can make costless announcements before choosing their strategies (e.g. Farrell, 1988), modifications of the assumption that players know that they choose their strategies independently (e.g. Colman & Bacharach, 1997; Krueger, 2008; Krueger & Acevedo, 2007), or special modes of reasoning (Bacharach, 1999, 2006; Sugden, 1993, 2005).

The cooperative value orientation offers a promising potential explanation for coordination, because it is interpreted in interdependence theory as the motive of Player $i$ to maximize the collective utility function defined as the sum of the objective payoffs to Players $i$ and $j$. However, it has three important limitations. First, it is defined for two-player games only, although the definition could be generalized to $n$-player games without difficulty. Second, in its simplest form it incorporates an implicit assumption that cooperatively motivated players assign equal weight to

[2] Player $i$ cannot solve the payoff dominance problem by applying the principle of insufficient reason, assuming that Player $j$ is equally likely to select $H_j$ or $L_j$, and concluding on this basis that $H_i$ is a best reply. In game theory, the players' rationality is assumed to be common knowledge, and it follows from this that, if Player $i$'s reasoning were valid, then $j$ would anticipate it and would play $H_j$, the best reply to $H_i$, *with certainty*. Player $i$ would anticipate this reply, refuting $i$'s initial assumption that the probability of $H_j$ was 1/2 and undermining the basis of $i$'s argument for choosing $H_i$. The same *reductio* proof works against any argument based on assigning subjective probabilities to $j$'s strategies (Colman, 2003; Colman, Pulford, & Rose, 2008).

their own and their co-players' objective payoffs. Whereas it seems reasonable to assume that players commonly assign *some* weight to their co-players' objective payoffs, the stronger assumption of *equal* weights seems unnecessarily restrictive. Third and most important, it turns out to be powerless to explain the payoff dominance phenomenon.

If both players are motivated by the cooperative value orientation, then, by definition,

$$U_i'(s_i, s_j) = f(u_i, u_j) = u_i + u_j,$$

for all $i, j \in N$. Fig. 5(b) shows the Stag Hunt game with these payoff transformations applied. The transformed payoff matrix still has Nash equilibria at $(C_1, C_2)$ and $(D_1, D_2)$, and both players still prefer $(C_1, C_2)$. Game theory provides no more reason for coordination in this transformed matrix than in the original (given) matrix: from a strategic point of view, nothing has changed. Fig. 5(d) shows that cooperative payoff transformations are of no help in solving the Hi-Lo game, leaving the strategic structure of the game unchanged. The cooperative value orientation, on its own, fails to explain coordination and the payoff dominance phenomenon, but with some crucial supplementary assumptions, team reasoning solves both problems.

Theories of team reasoning (Bacharach, 1999, 2006; Sugden, 1993, 2005) offer general solutions to common interest games and to the payoff dominance problem, even in instances such as the Hi-Lo game where Berge equilibrium surprisingly fails, and they also complement the Berge equilibrium by offering an alternative model of cooperation in the Prisoner's Dilemma and $n$-Player Prisoner's Dilemma games. However, they require a radical departure from orthodox game theory, assuming, as they do, not only that players are motivated to maximize collective payoffs, as in the cooperative social value orientation, but also that they adopt a distinctive mode of collective reasoning from preferences to decisions. In orthodox game-theoretic reasoning, players ask themselves: *What do I want, and what should I do to achieve it?* In team reasoning, they ask: *What do we want, and what should I do to help achieve it?* Even in the most challenging Hi-Lo game shown in Fig. 5(c), the answer is immediately obvious—*We want* $(H_1, H_2)$ *and, in my role as Player* $i$, *I should play my part in achieving it by choosing* $H_i$. Team-reasoning players begin by searching for a profile of strategies that maximizes the collective payoff of the group of players. If one profile Pareto-dominates all others, then that profile obviously maximizes the collective payoff. More generally, and in the spirit of the cooperative value orientation, if there exists a strategy profile that maximizes the sum of payoffs to the individual players,[3] then that profile is collectively optimal, even if it is not a Nash equilibrium. If the optimizing profile is unique, then team-reasoning players select and play their component strategies of it. If it is not unique, then the theory is indeterminate. Standard individual reasoning is subsumed within team reasoning as a special case in which the team is a singleton. There is evidence from experimental games that team reasoning occurs quite frequently in practice (Colman et al., 2008).

## 4. Related literature

Previous research into the Berge equilibrium and social value orientations has been sparse, and the relevant literature is widely scattered. The Berge equilibrium concept was introduced intuitively by the French mathematician Claude Berge (1957) for noncooperative games (chap. 1, Section 7) and coalition games (chap. 5, Section 27) at a time when he was a visiting professor at Princeton University and a colleague there of John Nash, who was on an Alfred P. Sloan grant. A brief review of Berge's book by Shubik (1961), published in *Econometrica*, could hardly have done more to discourage economists from reading it and may have contributed to its subsequent neglect in the English-speaking world: "The argument is presented in a highly abstract manner and no consideration is given to applications to economics.... The book will be of little direct interest to economists" (p. 821).

Berge's (1957) book was translated into Russian in 1961, and Zhukovskii (1985) later formalized the equilibrium in the context of differential games. Condition (2) was first suggested by Zhukovskii's doctoral student Konstantin Vaisman and adopted by Zhukovskii and Chikrii (1994, pp. 119–143) in a book published later in the same year, also in Russian. Research into the Berge equilibrium did not spread beyond the borders of the former USSR until Abalo and Kostreva (2004, 2005) published the first existence theorems for pure-strategy Berge equilibria in strategic-form games based on an existence theorem for differential games proposed by Radjef (1988). Radjef, who was a student of Zhukovskii, was one of many Algerian students trained in Russian universities, and that is why research published in Russian journals has been disseminated in some French journals by Algerian researchers. Nessah, Larbani, and Tazdaït (2007) and Larbani and Nessah (2008) showed the conditions set by Abalo and Kostreva to be insufficient to prove the existence of a pure-strategy Berge equilibria, and they proposed a new existence theorem. The Berge equilibrium has, up to now, received only a small amount of attention from mathematicians and virtually none from economists and other social and behavioral scientists.

The concept of social value orientations, interpreted in terms of payoff transformations, was introduced by McClintock (1972) and Messick and McClintock (1968), who focused attention initially on individualistic, cooperative, and competitive orientations. The idea was developed in interdependence theory by Kelley et al. (2003), Kelley and Thibaut (1978) and Thibaut and Kelley (1959). Research revealed that individual differences in baseline social value orientations are relatively stable over time (Kuhlman, Camac, & Cunha, 1986), correlate significantly with personality descriptions given by friends and roommates (Bem & Lord, 1979), and are predictive of activities in everyday life, such as volunteering for worthy causes (McClintock & Allison, 1989). A century earlier, Edgeworth (1881, pp. 102–104) had introduced the idea of other-regarding payoff transformations in a slightly more general form, in which Player $i$ maximizes the utility function $U'(s_i, s_j) = f_i(u_i, u_j) = pu_i + (1-p)u_j$, where $0 \leq p \leq 1$. Here, $p$ and $1-p$ represent relative weights that $i$ assigns to $u_i$ (own) and $u_j$ (co-player's) objective payoffs, respectively. Economists who have explored other-regarding utilities have generally reverted to this more general interpretation (e.g. Camerer, 1997, pp. 169–170; Fehr & Schmidt, 1999; Rabin, 1993).

## 5. Conclusions

Berge equilibrium provides a compelling model of cooperation in social dilemmas, including the Prisoner's Dilemma and $n$-Player Prisoner's Dilemma games. It deserves to be taken seriously, because an understanding of cooperation is an important goal, and especially because orthodox game theory has failed to explain experimental findings on cooperation adequately.

A Berge equilibrium occurs when players are mutually supportive in a game. It is thus reminiscent of the concepts of *solidarity* and *mutual aid* suggested by the Russian anarchist Peter Kropotkin (1902):

[3] In these circumstances, it makes no sense for players to weight their own and their co-players' payoffs unequally, because they are maximizing a collective payoff function.

> Love, sympathy and self-sacrifice certainly play an immense part in the progressive development of our moral feelings. But it is not love and not even sympathy upon which Society is based in mankind. It is the awareness[4] – be it only at the stage of an instinct – of human solidarity. It is the unconscious recognition of the force that is borrowed by each man from the practice of mutual aid; of the close dependency of every one's happiness upon the happiness of all; and of the sense of justice, or equity, which brings the individual to consider the rights of every other individual as equal to his own. (p. 5)

Berge equilibrium is a theoretical implication of the altruistic social value orientation, in which players have utility functions that motivate them to maximize each other's objective payoffs. Together with otherwise standard game-theoretic assumptions of rationality and common knowledge, this implies the emergence of Berge equilibria in games in which they exist. The existence of pure altruism has been debated by philosophers for centuries (see, e.g. Bentham, 1789; Comte, 1851/1875; Hume, 1739–1740), and there are skeptics who have argued that selfish egoism or individualism must be tautologically true, because people's actions reveal their individual preferences. Samuelson (1993), the originator of revealed preference theory, dismissed this view as mere casuistry:

> When the governess of infants caught in a burning building reenters it unobserved in a hopeless mission of rescue, casuists may argue; "She did it only to get the good feeling of doing it. Because otherwise she wouldn't have done it". Such argumentation (in Wolfgang Pauli's scathing phrase) is not even wrong. It is just boring, irrelevant, and in the technical sense of old-fashioned logical positivism "meaningless". (p. 143)

Pure altruism is evidently not the most common social value orientation in strategic interactions; but its existence cannot be ignored, and there are circumstances in which it seems entirely natural. Accumulating experimental evidence, reviewed in the Introduction (Section 1), makes it difficult to deny that there are circumstances in which altruism is a regular occurrence. In our benchmark game outlined informally in the Introduction (Section 1), altruism and Berge equilibrium seem intuitively compelling.

We have shown how the Berge equilibrium provides a natural model of cooperation in $n$-Player Prisoner's Dilemmas, and how it can also model coordination in some common interest games, although it is inadequate in the sharpest challenge, the pure coordination Hi-Lo matching game. The cooperative social value orientation, in which players are motivated to maximize their collective payoff, explains cooperation in social dilemmas and might be expected to explain coordination in common interest games. However, it turns out to be inadequate to explain coordination without additional theoretical apparatus provided by theories of team reasoning.

Berge equilibrium and team reasoning are both powerful theories of mutual support in games. They model two of the most important forms of social collaboration, namely cooperation in social dilemmas and coordination in common interest games. Team reasoning has received at least some attention from researchers in the social and behavioral sciences; Berge equilibrium has hitherto received virtually none, but our preliminary exploration suggests that it merits further investigation.

## References


Abalo, K., & Kostreva, M. (2004). Some existence theorems of Nash and Berge equilibria. *Applied Mathematics Letters*, *17*, 569–573.

Abalo, K., & Kostreva, M. (2005). Berge equilibrium: some results from fixed-point theorems. *Applied Mathematics and Computation*, *169*, 624–638.

Arnsperger, C., & Varoufakis, Y. (2003). Toward a theory of solidarity. *Erkenntnis*, *59*, 157–188.

Aumann, R. J., & Sorin, S. (1989). Cooperation and bounded recall. *Games and Economic Behavior*, *1*, 5–39.

Bacharach, M. (1999). Interactive team reasoning: a contribution to the theory of co-operation. *Research in Economics*, *53*, 117–147.

Bacharach, M. (2006). In N. Gold, & R. Sugden (Eds.), *Beyond individual choice: teams and frames in game theory*. Princeton, NJ: Princeton University Press.

Batson, C. D., & Ahmad, N. (2001). Empathy-induced altruism in a Prisoner's dilemma II: what if the target of empathy has defected? *European Journal of Social Psychology*, *31*, 25–36.

Batson, C. D., & Shaw, L. L. (1991). Evidence for altruism: toward a pluralism of prosocial motives. *Psychological Inquiry*, *2*, 107–122.

Bem, D. J., & Lord, C. G. (1979). Template matching: a proposal for probing the ecological validity of experimental settings in social psychology. *Journal of Personality and Social Psychology*, *37*, 833–846.

Bentham, J. (1789). *An introduction to the principles of morals and legislation*. Oxford: Clarendon Press.

Berge, C. (1957). *Théorie générale des jeux à n personnes* [General theory of $n$-person games]. Paris: Gauthier-Villars.

Bolton, G. E., & Ockenfels, A. (2000). ERC: a theory of equity, reciprocity, and competition. *American Economic Review*, *90*, 166–193.

Brosnan, S. F., & de Waal, F. B. M. (2002). A proximate perspective on reciprocal altruism. *Human Nature*, *13*, 129–152.

Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of Economic Perspectives*, *11*, 167–188.

Camerer, C. F., & Fehr, E. (2006). When does economic man dominate social behavior? *Science*, *311*, 47–52.

Camerer, C. F., & Thaler, R. H. (1995). Anomalies: ultimatums, dictators, and manners. *Journal of Economic Perspectives*, *9*, 209–219.

Colman, A. M. (1995). *Game theory and its applications in the social and biological sciences* (2nd ed.) Oxford: Butterworth-Heinemann & London, Routledge.

Colman, A. M. (2003). Cooperation, psychological game theory, and limitations of rationality in social interaction. *Behavioral and Brain Sciences*, *26*, 139–153.

Colman, A. M., & Bacharach, M. (1997). Payoff dominance and the Stackelberg heuristic. *Theory and Decision*, *43*, 1–19.

Colman, A. M., Pulford, B. D., & Rose, J. (2008). Collective rationality in interactive decisions: evidence for team reasoning. *Acta Psychologica*, *128*, 387–397.

Comte, 1. A. (1875). *Trans. J. Bridges*: *Vol. 1. System of positive polity*. London: Longmans, Green, Original French work published 1851.

Edgeworth, F. Y. (1881). *Mathematical psychics: an essay on the application of mathematics to the moral sciences*. London: Kegan Paul.

Farrell, J. (1988). Communication, coordination and Nash equilibrium. *Economics Letters*, *27*, 209–214.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, *114*, 817–868.

Gibbard, A. (1974). A Pareto-consistent libertarian claim. *Journal of Economic Theory*, *7*, 388–410.

Hamburger, H. (1973). $N$-person Prisoner's dilemma. *Journal of Mathematical Sociology*, *3*, 27–48.

Harsanyi, J. C., & Selten, R. (1988). *A general theory of equilibrium selection in games*. Cambridge, MA: MIT Press.

Hume, D. (1739–1740). *A treatise of human nature: being an attempt to introduce the experimental method of reasoning into moral subjects*. London: Thomas Longman.

Kelley, H. H., Holmes, J. G., Kerr, N. L., Reis, H. T., Rusbult, C. E., & Van Lange, P. A. M. (2003). *An atlas of interpersonal situations*. Cambridge: Cambridge University Press.

Kelley, H. H., & Thibaut, J. W. (1978). *Interpersonal relations: a theory of interdependence*. New York: Wiley.

Kropotkin, P. (1902). *Mutual aid: a factor of evolution*. New York: Doubleday.

Krueger, J. I. (2008). Methodological individualism in experimental games: not so easily dismissed. *Acta Psychologica*, *128*, 398–401.

Krueger, J. I., & Acevedo, M. (2007). Perceptions of self and other in the Prisoner's dilemma: outcome bias and evidential reasoning. *American Journal of Psychology*, *120*, 593–618.

Kuhlman, D. M., Camac, C., & Cunha, D. A. (1986). Individual differences in social orientation. In H. Wilke, D. Messick, & C. Rutte (Eds.), *Experimental social dilemmas* (pp. 151–176). New York: Verlag Peter Lang.

Larbani, M., & Nessah, R. (2008). A note on the existence of Berge and Berge–Nash equilibria. *Mathematical Social Sciences*, *55*, 258–271.

Lewis, D. K. (1969). *Convention: a philosophical study*. Cambridge, MA: Harvard University Press.

Marler, P. (1955). Characteristics of some animal calls. *Nature*, *176*, 6–8.

Marler, P. (1959). Developments in the study of animal communication. In P. R. Bell (Ed.), *Darwin's biological work* (pp. 150–206). Cambridge: Cambridge University Press, 329–335.

May, R. (2006). Threats to tomorrow's world: address of the President, Lord May of Oxford OM AC FRS, given at the anniversary meeting on 30 November 2005. *Notes and Records of the Royal Society*, *60*, 109–130.

Maynard Smith, J. (1965). The evolution of alarm calls. *American Naturalist*, *99*, 59–63.

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*, 15–18.

McClintock, C. G. (1972). Social motivation: a set of propositions. *Behavioral Science*, *17*, 438–454.

McClintock, C. G., & Allison, S. T. (1989). Social value orientation and helping behavior. *Journal of Applied Social Psychology*, *19*, 353–362.


[4] The original has "conscience"—obviously an error by an author whose native tongue was not English.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*, 1–25.
Nash, J. F. (1950). Equilibrium points in *n*-person games. *Proceedings of the National Academy of Sciences, USA*, *36*, 48–49.
Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, *54*, 286–295.
Nessah, R., Larbani, M., & Tazdaït, T. (2007). A note on Berge equilibrium. *Applied Mathematics Letters*, *20*, 926–932.
Parco, J. E., Rapoport, A., & Stein, W. E. (2002). Effects of financial incentives on the breakdown of mutual trust. *Psychological Science*, *13*, 292–297.
Rabin, M. (1993). Incorporating fairness into game theory and economics. *American Economic Review*, *83*, 1281–1302.
Radjef, M. S. (1998). Sur l'existence d'un équilibre de Berge pour un jeu différentiel à *n* personnes. [On the existence of a Berge equilibrium for an *n*-person differential game]. *Cahiers Mathématiques de l'Université d'Oran*, *1*, 89–93.
Rapoport, A. (1967). Exploiter, Leader, Hero, and Martyr: the four archetypes of the $2 \times 2$ game. *Behavioral Science*, *12*, 81–84.
Rapoport, A., Guyer, M. J., & Gordon, D. G. (1976). *The* $2 \times 2$ *game*. Ann Arbor, MI: University of Michigan Press.
Rousseau, J. -J. (1755). Discours sur l'origine et les fondements de l'inégalité parmi les hommes. [Discourse on the origin and the foundations of inequality among men]. In J.-J. Rousseau, (Ed), *Oeuvres complètes. Vol. 3. Edition Pléiade*. Paris.
Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology*, *54*, 351–375.
Russell, B. (1959). *Common sense and nuclear warfare*. London: Allen & Unwin.
Sally, D. (1995). Conversation and cooperation: a meta-analysis of experiments from 1958 to 1992. *Rationality and Society*, *7*, 58–92.
Samuelson, P. A. (1993). Altruism as a problem involving group versus individual selection in economics and biology. *American Economic Review*, *83*, 143–148.
Schelling, T. C. (1973). Hockey helmets, concealed weapons, and daylight saving: a study of binary choices with externalities. *Journal of Conflict Resolution*, *17*, 381–428.
Sen, A. K. (1969). A game-theoretic analysis of theories of collectivism in allocation. In T. Majumdar (Ed.), *Growth and choice: essays in honour of U.N. Ghosal* (pp. 1–17). Calcutta: Oxford University Press.
Shubik, M. (1961). Review of C. Berge, general theory of *n*-person games. *Econometrica*, *29*, 821.
Sugden, R. (1986). *The economics of rights, cooperation and welfare*. Blackwell.
Sugden, R. (1993). Thinking as a team: towards an explanation of nonselfish behaviour. *Social Philosophy and Policy*, *10*, 69–89.
Sugden, R. (2005). The logic of team reasoning. In N. Gold (Ed.), *Teamwork: multi-disciplinary perspectives* (pp. 181–199). Basingstoke: Palgrave Macmillan.
Suleiman, R., Budescu, D. V., Fischer, I., & Messick, D. M. (2004). *Contemporary psychological research on social dilemmas*. Cambridge: Cambridge University Press.
Thibaut, J. W., & Kelley, H. H. (1959). *The social psychology of groups*. New York: Wiley.
Trivers, R. (2005). Reciprocal altruism: 30 years later. In C. P. van Schaik, & P. M. Kappeler (Eds.), *Cooperation in primates and humans: mechanisms and evolution* (pp. 67–83). Berlin: Springer-Verlag.
Tucker, A. (1950/2001). A two-person dilemma. Stanford University. In E. Rasmussen (Ed.), *Readings in games and information* (pp. 7–8). Malden, MA: Blackwell (Reprinted) (unpublished notes).
Van Lange, P. A. M. (2000). *European Review of Social Psychology*: *11*. *Beyond self-interest: a set of propositions relevant to interpersonal orientations* (pp. 297–330). London: Wiley.
Wilson, D. R., & Evans, C. S. (2008). Mating success increases alarm-calling effort in male fowl, Gallus gallus. *Animal Behavior*, *76*, 2029–2035.
Zhukovskii, V. I. (1985). In P. Kenderov (Ed.), *Mathematical methods in operations research*, *Matematiceskie metody v issledovanii operacij* [Some problems of non-antagonistic differential games] (pp. 103–195). Sofia: Bulgarian Academy of Sciences.
Zhukovskii, V. I., & Chikrii, A. A. (1994). *Lineino-kvadratichnye differentsial'nye igry* [Linear-quadratic differential games]. Kiev: Naoukova Doumka.