View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by Springer - Publisher Connector

Auton Agent Multi-Agent Syst (2012) 25:284–312
DOI 10.1007/s10458-011-9176-3

Dependence in games and dependence games

Davide Grossi · Paolo Turrini

Published online: 24 June 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract In the multi-agent systems community, dependence theory and game theory are often presented as two alternative perspectives on the analysis of agent interaction. The paper presents a formal analysis of a notion of dependence between players, given in terms of standard game-theoretic notions of rationality such as dominant strategy and best response. This brings the notion of dependence within the realm of game theory providing it with the sort of mathematical foundations which still lacks. Concretely, the paper presents two results: first, it shows how the proposed notion of dependence allows for an elegant characterization of a property of reciprocity for outcomes in strategic games; and second, it shows how the notion can be used to define new classes of coalitional games, where coalitions can force outcomes only in the presence of reciprocal dependencies.

Keywords Dependence theory · Game theory · Multi-agent systems

1 Introduction

Dependence theory was born within the social sciences [7] and brought into multi-agent systems (henceforth, MAS) by [5,6] where, in the last decade, its ideas have made their way into several lines of research (e.g., [1,3,17,18]). Dependence theory has, at the moment, several versions and no unified theory. Yet, the aim of the theory is clear:

> "One of the fundamental notions of social interaction is the *dependence* relation among agents. In our opinion, the terminology for describing interaction in a multi-agent world is necessarily based on an analytic description of this relation. Starting from such a

D. Grossi
Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, The Netherlands
e-mail: d.grossi@uva.nl

P. Turrini (✉)
Department of Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands
e-mail: paolo@cs.uu.nl

terminology, it is possible to devise a calculus to obtain predictions and make choices that simulate human behavior" [6, p. 2].

Therefore, dependence theory addresses two issues: (i) the representation of dependence relations among the agents in a system (to the nature of this relation we will come back in Sect. 3); (ii) the use of such information as a means to obtain predictions about the behavior of the system. While all contributions to dependence theory have focused on the first point, the second challenge, "devise a calculus to obtain predictions", has been mainly left to computer simulation [17].

The present paper takes up these two challenges outlining a theory of dependence in strategic interaction which makes use of standard game-theoretic notions and techniques. The theory we propose builds up on a notion of dependence relations in strategic games, inspired from informal literature on dependence theory in MAS (e.g., [5,6]):

Player i depends on player j for reaching outcome s, within a given game, if and only if j plays a strategy, in the profile determining s, which is a best response (or a dominant strategy) for i.

The aim of the paper is to provide a thorough analysis of the above informal definition, which will be made formal in Definition 9. Concretely, the paper presents two results. First, it shows that this notion of dependence allows for the characterization of an original notion of reciprocity for strategic games (Theorem 1), i.e. how players in a strategic interaction can profitably make use of their dependence relation. Second, it shows that the notion of dependence can be fruitfully applied to ground coalition formation. The class of coalitional games where coalitions can force outcomes only in the presence of reciprocity —here called *dependence games*—can be directly linked with standard solution concepts used in cooperative game theory (Theorems 2, 3). The paper generalizes results presented in [9], discusses further examples, and puts the results in perspective with game theory and with related work in MAS. Definitions and results will be thoroughly illustrated by examples and, in particular, by discussing their application to two- and three-person games.

As to the motivation behind the research presented here, it can be said that the paper moves from the authors' impression that, within the MAS community, dependence theory and game theory are erroneously considered to be alternative, when not incompatible, paradigms for the analysis of social interaction. It is our conviction that the theory of games and that of dependence are highly compatible endeavors which can highly benefit from one another in the field of MAS. On the one hand dependence theory can be incorporated into the highly developed mathematical framework of game theory, obtaining the sort of foundations that still lacks. On the other hand, game theory can incorporate a dependence-theoretic perspective on the analysis of strategic interaction.

Outline of the paper

Section 2 briefly introduces the basic notions of game theory with which we will work in the rest of the paper. Section 3 introduces our formal analysis of the concept of dependence, relating it to informal definitions available in the literature on dependence theory. It then moves on to the characterization of reciprocal outcomes in games. Starting from the notion of reciprocity, Section 4 introduces a notion of agreement among players in a game, and on this ground it defines and studies a specific class of coalitional games. Section 5 describes in detail some further examples of our analysis proposing also an application of the framework

Fig. 1 Examples of two player strategic games in game matrices. Ordinal preferences are represented, as usual, by means of numerical payoffs. U and D denote the strategies 'up' and, respectively, 'down' for the row player. L and R denote the strategies '*left*' and '*right*' for the column player

	L	R
U	2,2	0,3
D	3,0	1,1

Prisoner's dilemma

	L	R
U	3,3	2,2
D	2,2	1,1

Full Convergence

	L	R
U	1,1	0,0
D	0,0	1,1

Coordination

	L	R
U	3,3	2,2
D	2,5	1,1

Partial Convergence

to transferable utility games. Finally, Sect. 6 discusses related and future work, and concludes the paper.

2 Preliminaries: game theory

The present section introduces the basic game-theoretic notions used in the paper. All definitions will be based on an ordinal notion of preference. Main sources for this preliminary section are [12,13].

2.1 Strategic games, sub-games and solution concepts

Let us start with the definition of a game in strategic form.

Definition 1 (*Game*) A (strategic form) game is a tuple $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$[1] where:

- N is a non-empty set of players;
- S is a non-empty set of outcomes;
- Σ_i is a non-empty set of strategies for each player $i \in N$;
- $\succeq_i$ is a total preorder on S;
- $o : \times_{i \in N} \Sigma_i \to S$ is a function from the set $\times_{i \in N} \Sigma_i$ of strategy profiles to the set of outcomes S.[2]

Examples of games represented as payoff matrices are given in Fig. 1. The terms 'agent' and 'player' will be used interchangeably. Before we continue with the next definition, let us first introduce some further notation. Strategy profiles will be denoted $\sigma, \sigma' \ldots$. Given a strategy profile σ and a player i, σ_i denotes the strategy chosen by i in σ, i.e., the ith projection of σ, and σ_{-i} denotes the profile consisting of all the strategies of the players except i. So, a profile σ can be seen as the juxtaposition of σ_i and σ_{-i}, in symbols, $\sigma = (\sigma_i, \sigma_{-i})$. More generally, σ_C, for a coalition $C \subseteq N$, denotes the tuple of strategies performed by the agents in C in profile σ, i.e. an element of $\times_{i \in C} \Sigma_i$. Given a coalition C, in order to denote the set of agents not belonging to C we use notation $\overline{C}$. We now define the notion of sub-game of a game in strategic form.

Definition 2 (*Sub-game*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, σ be a strategy profile, and $C \subseteq N$. The subgame of $\mathcal{G}$ defined by σ_C is a game $\mathcal{G} \downarrow \sigma_c = (N', S', \Sigma'_i, \succeq'_i, o')$ such that:

[1] In order to streamline the notation, we assume here the convention, also assumed in [13], to denote as Σ_i the set of all strategy spaces $\{\Sigma_i\}_{i \in N}$. The same we do for the set of preferences.

[2] The outcome function is often assumed to be a bijection and it is consequently dispensed with [13]. However in the field of social choice theory several contributions exist where bijectivity is considered a restrictive constraint [12]. As different scenarios may require different assumptions, we study the outcome function in its most general form.

- $N' = \overline{C}$;
- $S' = S - \{s \mid \exists \sigma' \text{ S.T. } s = o(\sigma') \text{ AND } \sigma'_C \neq \sigma_C\}$;
- for all $i \in \overline{C}$, $\Sigma'_i = \Sigma_i$;
- for all $i \in \overline{C}$, $\succeq'_i = \succeq_i \cap S' \times S'$;
- $o' : \times_{i \in N'} \Sigma_i \to S'$ is a function from the strategy profiles to outcomes such that for all $\sigma' \in \times_{i \in N'} \Sigma_i$, $o'(\sigma') = o(\sigma', \sigma_C)$.

A sub-game $\mathcal{G} \downarrow \sigma_C$ of $\mathcal{G}$ is obtained from $\mathcal{G}$ once the coalitional strategy σ_C of the set of players in C is fixed. Intuitively, it as a snapshot of what is still 'left to play' once the players in C have made their choice. Therefore, in a sub-game, the only players actually playing are the ones outside C (first item in the definition); the available outcomes are only those that can be reached once the strategies of C are fixed (second item); the strategies available to the players remain the same (third item); players' preferences are restricted to the available outcomes (fourth item); and the outcome function corresponds to the one of the original game for that part of profiles which are available in the sub-game (fifth item). It is worth stressing that a sub-game is still a game in the sense of Definition 1.

As to the solution concepts, we will work with the Nash equilibrium, which we will typically refer to as best response equilibrium (BR-equilibrium), and the dominant strategy equilibrium (DS-equilibrium). Let us first introduce the notions of best response and dominant strategy.

Definition 3 (*Best response and dominant strategy*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, and σ a profile. We say that:

- strategy σ_i is a *best response* for i in σ iff $\forall \sigma'_i \in \Sigma_i : o(\sigma) \succeq_i o(\sigma'_i, \sigma_{-i})$;
- strategy σ_i is a *dominant strategy* for i iff $\forall \sigma' \in \times_{i \in N} \Sigma_i : o(\sigma_i, \sigma'_{-i}) \succeq_i o(\sigma')$.

In words, a strategy of i is a best response for i, in a given profile, whenever i has no strategy available that would lead to a strictly better outcome for him/herself, provided the rest of the players stick to their choices. The notion of dominant strategy is stronger. A strategy is a dominant strategy for i whenever no other choice of i would guarantee him/her a strictly better outcome, no matter what the rest of the players chooses to do. It is easy to see that a dominant strategy is a best response under any strategy profile.

With these auxiliary notions in place, we can now introduce the two solution concepts of interest for the paper.

Definition 4 (*Equilibria*) Let $\mathcal{G}$ be a game. A strategy profile σ is:

- a *BR-equilibrium* (Nash equilibrium) if σ_i is a best response in σ for all players i in N;
- a *DS-equilibrium* (dominant strategy equilibrium) if σ_i is a dominant strategy for all players i in N.

In words, a Nash or BR-equilibrium is a profile in which no player can do strictly better for him/herself by deviating from his/her current strategy, i.e., by changing his/her strategy while assuming the other players stick to theirs. A DS-equilibrium is a stronger solution concept. It is a profile in which all players choose a strategy which is preferable no matter what all the other players choose.

Example 1 (*Solution concepts*) Consider the four games given in Fig. 1. We will illustrate Definition 4 by looking at their equilibria. In the prisoner's dilemma the profile (D, R)—the profile commonly known as 'defect-defect' profile—is both a BR and a DS-equilibrium. The same is the case for profile (U, L) in the full convergence and partial convergence games.

The coordination game has two BR-equilibria, namely (U, L) and (D, R), but no DS-equilibrium as players do not have a strategy which under any circumstance would lead to a preferable outcome.

2.2 Coalitional games and the core

In addition to the games in strategic form (Definition 1) we will also work with cooperative, or coalitional, games. Unlike the games in strategic form, coalitional games, which have been extensively studied already in [19], model interactive situations by explicitly representing the possibility of cooperation among coalitions of agents. In particular, we will work with coalitional games with non-transferable payoffs [13], that is, with ordinal versions of coalitional games. These are represented abstractly by means of so-called effectivity functions [12].

Definition 5 (*Coalitional game*) A coalitional game is a tuple $\mathcal{C} = (N, S, E, \succeq_i)$ where:

- N is a non-empty set of players;
- S is a non-empty set of outcomes;
- E is function $E : 2^N \to 2^{2^S}$ (the *effectivity function*) from sets of players to sets of sets of outcomes;
- $\succeq_i$ is a total preorder on S.

Just like games in strategic form, coalitional games consist of a set of players endowed with preferences over a set of outcomes. But unlike games in strategic form, interaction is here modeled via the effectivity function. An effectivity function associates to a coalition a set of sets of outcomes and the fact that a set of outcomes X belongs to the effectivity $E(C)$ of C is usually understood as the coalition C being able to force the interaction to end up in an outcome in X.

In this paper we are interested in coalitional games only in as much as they provide a representation of games in strategic form which explicitly allows for cooperation. In other words, we use coalitional games only to describe the cooperative possibilities of players in strategic interaction. We now recall how to obtain a coalitional game from a strategic one (cf. [12]). These games in particular will be the object of study in Sect. 4.

Definition 6 (*Coalitional games from strategic ones*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a strategic game. The coalitional game $\mathcal{C}^{\mathcal{G}} = (N, S, E^{\mathcal{G}}, \succeq_i)$ of $\mathcal{G}$ is a coalitional game where the effectivity function $E^{\mathcal{G}}$ is defined as follows:

$$X \in E^{\mathcal{G}}(C) \Leftrightarrow \exists \sigma_C \forall \sigma_{\overline{C}}\, o(\sigma_C, \sigma_{\overline{C}}) \in X.$$

So the coalitional game $\mathcal{C}$ of a given game $\mathcal{G}$ in strategic form consists of $\mathcal{G}$'s set of agents, $\mathcal{G}$'s set of outcomes and $\mathcal{G}$'s agents' preferences. The effectivity function $E^{\mathcal{G}}$ is then defined in such a way that, for any coalition C, $E^{\mathcal{G}}(C)$ consists of all those sets of outcomes within which C can force the game $\mathcal{G}$ to end up, no matter what the rest of the players do.

Example 2 (*Effectivity function of the prisoner's dilemma*) Consider again the prisoner's dilemma (Fig. 1). From a cooperative point of view there are four coalitions available: $\{column\}$, $\{row\}$, $\{column, row\}$ and the empty coalition $\emptyset$. Let us abbreviate the set $\{(U, L), (U, R), (D, L), (D, R)\}$ with W. By applying Definition 6 we obtain the following effectivity function E:

$$
\begin{aligned}
E(\emptyset) &= \{W\} \\
E(\{column\}) &= \{X \subseteq W \mid \{(U, L), (D, L)\} \subseteq X \quad \text{or} \quad \{(U, R), (D, R)\} \subseteq X\} \\
E(\{row\}) &= \{X \subseteq W \mid \{(U, L), (U, R)\} \subseteq X \quad \text{or} \quad \{(D, L), (D, R)\} \subseteq X\} \\
E(\{column, row\}) &= \{X \subseteq W \mid X \neq \emptyset\}
\end{aligned}
$$

Intuitively, as the coalition grows, so grow its possibilities. Trivially, the empty coalition has no way of 'restraining' the game to any particular set of outcomes. Each player can either force the game to end up in the sets of outcomes consisting of the rows of the game matrix—by choosing U or D—or, respectively, the ones consisting of the columns of the game matrix—by choosing L or R. The coalition consisting of both players can not only force the game to end up in the sets belonging to the effectivity function of each player, but it can also force the game towards any of the four singleton sets. In other words, by coordinating their actions, the two players together can decide which outcome should result from the game.

Finally, we consider the most common solution concept for coalitional games, the core. This is inspired by the same sort of rationale behind the notion of Nash equilibrium in games in strategic form: an outcome is stable—and hence belongs to the core of the game—if no coalition can profitably deviate obtaining an outcome which is better for all the members of the coalition.

Definition 7 (*The Core*) Let $\mathcal{C} = (N, S, E, \succeq_i)$ be a coalitional game. We say that a state $s \in S$ is dominated in $\mathcal{C}$ if for some C and $X \in E(C)$ it holds that $x \succ_i s$ for all $x \in X$, $i \in C$. The core of $\mathcal{C}$, in symbols $CORE(\mathcal{C})$ is the set of undominated states.

As an illustration, we show how the notion of core rationalizes the cooperative solution of the prisoner's dilemma.

Example 3 (*The core of the coalitional version of the prisoner's dilemma*) Recall the effectivity function E described in Example 2. The coalitional version of the prisoner's dilemma scenario is then a coalitional game consisting of the two players *column* and *row*, the four set of outcomes $\{U, D\} \times \{L, R\}$, the effectivity function E, and the preferences given in the top left game matrix in Fig. 1. The core of such game is the set $\{(U, L)\}$ in which the two players cooperate. This is so because neither can the coalition $\{column, row\}$ guarantee a better outcome for both players, nor can either of the players guarantee outcomes that are all strictly better than (U, L). To appreciate this notice, for example, that none of the sets of outcomes in $E(\{column\})$ (see Example 2) consists of outcomes that are all strictly preferred to (U, L) by *column*.

We have now concluded the introduction of the game-theoretic machinery which will be used in the rest of the paper, and can move to the presentation of our formal theory of dependence.

3 Dependence in games

Dependence theory, as developed within artificial intelligence and MAS, has been mainly inspired by work in the social sciences such as [7]. It moves from presuppositions that are clearly shared by the theory of games—eminently the fact that the outcome of social interaction depends on the choices of different agents—but it emphasizes, rather than the strategic aspect of agents' choices, the interdependencies existing between them in terms of what they want and what they choose:

> "Sociality obviously presupposes two or more players in a common shared world. A 'Common World' implies that there is interference among the actions and goals of the players: the effects of the action of one player are relevant for the goals of another: i.e., they either favor the achievement or maintenance of some goals of the other's (positive interference), or threat some of them (negative interference)." [5, pp. 161–162]

In this view, what underpins the analysis of social interaction is the idea that agents can favor or hinder each other's goals.

The present section shows how, by tweaking some basic game-theoretic notions, this perspective on social interaction can be accommodated within the theory of games. The section proceeds with a formal definition and analysis of the notion of dependence in games.

3.1 Dependence relations

The literature on dependence theory deals with a number of different relations of dependence. Yet, in its most essential form, a dependence relation is a relation occurring between two agents i and j with respect to a certain state (or goal) which i wants to achieve but which it cannot achieve without some appropriate action of j.

> "x depends on y with regard to an act useful for realizing a state p when p is a goal of x's and x is unable to realize p while y is able to do so." [6, p. 4]

By taking a game theoretic perspective, this informal relation acquires a precise and natural meaning: a player i depends on a player j for the realization of a state p, i.e., of the strategy profile σ such that $o(\sigma) = p$, when, in order for σ to occur, j has to favor i, that is, it has to play in i's interest. To put it otherwise, i depends on j for σ when, in order to achieve σ, j has to do a favor to i by playing σ_j (which is obviously not under i's control).[3]

The latter intuition is made clear as follows.

Definition 8 (*Best for someone else*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and σ a profile, and let $i, j \in N$. We say that:

- strategy σ_j is a *best response for* i in σ iff $\forall \sigma'_j \in \Sigma_j : o(\sigma) \succeq_i o(\sigma'_j, \sigma_{-j})$;
- strategy σ_j is a *dominant strategy for* i iff $\forall \sigma' \in \times_{i \in N} \Sigma_i : o(\sigma_i, \sigma'_{-i}) \succeq_i o(\sigma')$.

It is easy to see that Definition 8 is a simple generalization of the definitions of best response and dominant strategy given in Definition 3. We simply allow the player holding the preference to be different from the player whose strategies are considered. We take this to be natural formalizations, within a standard game theoretic setting, of notions such as "doing a favor to" or "playing in the interest of".

On the base of Definition 8 we can formulate a precise game-theoretic notion of dependence in the following way.

Definition 9 (*Dependence*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and let $i, j \in N$. We say that:

[3] It might be worth noticing that while the notion of dependence relation we use is a three-place one, the literature on dependence theory often considers richer relations with higher arity, incorporating ingredients such as plans and actions (see, for instance [18]).

1. i *BR-depends* on j in profile σ iff σ_j is a best response for i in σ;
2. i *DS-depends* on j in profile σ iff σ_j is a dominant strategy for i.

That i BR-depends on j for σ will often be denoted $i\,R_\sigma^{BR}\,j$. Similarly, that i DS-depends on j for σ is often schematically denoted $i\,R_\sigma^{DS}\,j$.

Intuitively, i depends on j for profile σ in a best response sense if, in σ, j plays a strategy which is a best response for i given the strategies in σ_{-j} (and hence given the choice of i itself). It depends in a dominant strategy sense if, in σ, j plays a dominant strategy for i.

Definition 9 deserves a few remarks. The first thing to notice is that the notion of dependence arising from the definition is based on an underlying notion of rationality. In our case we opted for the ones that are, arguably, most standard in a pure strategy setting like ours: best response and dominant strategy. But it must be clear that other choices are possible (e.g. strict best response, strict dominance) and that Definition 9 could be easily extended to accommodate them.

Secondly, a consequence of the definition is that, given any game, we can always associate to any profile σ a binary relation—R_σ^{BR} or R_σ^{DS}—on the set of players which describes who depends on whom for the realization of that profile. In other words, we can associate to each profile σ a graph $\langle N, R_\sigma^{BR}\rangle$, or a graph $\langle N, R_\sigma^{DS}\rangle$, which provide a structural description of the sort of dependencies at work in the underlying game. We call these graphs *dependence graphs*.

Example 4 (*The dependence graph of the prisoner's dilemma*) Consider again the prisoner's dilemma in Fig. 1. Its best response dependence graph is depicted in Fig. 2. There we notice that, for instance, the relation $R_{(U,R)}^{BR}$ depicted in the up-right corner is such that *column* depends on him/herself (it is a reflexive point), as it plays its own best response, but also on *row*, as *row* does not play its own best response but a best response for *column*. Therefore (U, R) displays some kind of 'unbalance'. On the contrary, the graph associated to (U, L) depicts a cycle of BR-dependence in which *row* plays a best response for *column* and, vice versa, *column* for *row*.

The last remark worth making is that, in general, relations R_σ^{BR} and R_σ^{DS} do not enjoy any particular structural property. However, when they do, such structural properties can have a precise game theoretic meaning. The following simple fact gives a simple example of how structural properties of dependence graphs relate to game-theoretic properties of the underlying games.

Fact 1 (Reflexive dependencies and equilibria) *Let $\mathcal{G}$ be a game and let $x \in \{BR, DS\}$. It holds that: for any profile σ, R_σ^x is reflexive iff σ is an x-equilibrium.*

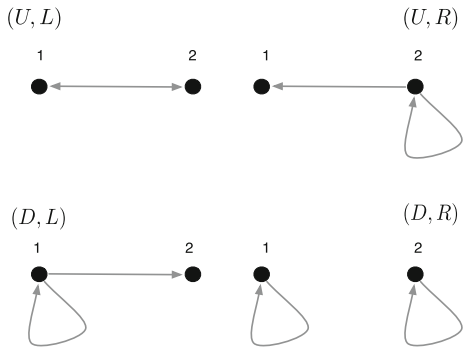

Fig. 2 The BR-dependence graph of the prisoner's dilemma: 1 and 2 denote player *row* and, respectively, player *column*

Proof Two claims must be proven. [First claim for $x = BR$] From left to right, we assume that $\forall i \in N, i R_\sigma^{BR} i$. From Definition 9, it follows that $\forall i \in N, \forall \sigma' : o(\sigma) \succeq_i o(\sigma_i', \sigma_{-i})$, that is, σ is a Nash equilibrium. From left to right, we assume that σ is a Nash equilibrium. From this it follows that $\forall i \in N \sigma_i$ is a best response for i, from which the reflexivity of R_σ^{BR} follows by Definition 9. [Second claim for $x = DS$] An analogous proof applies. □

In other words, any profile in which players depend on themselves—either in a best response or in a dominant strategy sense—is an equilibrium of the corresponding type—BR or DS. Figure 2 offers a good pictorial example. The 'defect-defect' profile (D, R), the Nash equilibrium, gives indeed rise to a BR-dependence relation which is reflexive.

3.2 Cycles

We have seen above how the reflexivity of dependence is related to the existence of equilibria (Fact 1). In this section we move to a more general property of dependence relations, the existence of cycles. The literature on dependence theory in MAS puts particular emphasis on this property as cycles intuitively suggest that there exist common ground for cooperation: if an individual depends on an other individual, and the latter depends in turn on the first to achieve a specific outcome, the choice of that outcome means that the individuals are doing each other a favor. This perspective is very clearly expressed, for instance, in [1,3], where dependence cycles are taken to signal the possibility of social interaction between players of a *do-ut-des* (give-to-get) type.

In that literature, however, dependence relations are considered as given—they do not arise from underlying structures such as games—and so are cycles, whose importance is not motivated in terms of some underlying rationale, but is taken for granted. In this and the following sections (Sects. 3.2–3.4) we show how, starting from dependence relations that arise from an underlying game (Definition 9), we can give precise game-theoretic reasons for the significance of dependence cycles in strategic settings. So let us start with a definition of what a dependence cycle is.

Definition 10 (*Dependence cycles*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, (N, R_σ^x) be its dependence structure for profile σ with $x \in \{BR, DS\}$, and let $i, j \in N$. An R_σ^x-dependence cycle $\mathfrak{c}$ of length $k-1$ in $\mathcal{G}$ is a tuple $(a_1, \ldots, a_k)$ such that:

1. $a_1, \ldots, a_k \in N$;
2. $a_1 = a_k$;
3. $\forall a_i, a_j$ with $1 \leq i \neq j < k, a_i \neq a_j$;
4. $a_1 R_\sigma^x a_2 R_\sigma^x \ldots R_\sigma^x a_{k-1} R_\sigma^x a_k$.

Given a cycle $\mathfrak{c} = (a_1, \ldots, a_k)$, its orbit $O(\mathfrak{c}) = \{a_1, \ldots, a_{k-1}\}$ denotes the set of its elements.

In other words, cycles are sequences of pairwise different players, except for the first and the last which are equal, such that all players are linked by a dependence relation. Note that the definition allows for cycles of length 1, whose orbit is a singleton, i.e., reflexive arcs. Those are the cycles occurring at reflexive points in the graph.

We have already seen in Example 4 that the cooperative outcome of the prisoner's dilemma exhibits a cycle linking *row* and *column* (see Fig. 2). Even more interesting are cycles in games with more than two players.

Fig. 3 A three person game. Player 1 denotes *row*, player 2 *column*, and player 3 chooses between the *right* and *left* matrices

	g	$\neg g$
g	3, 3, 3	2, 4, 2
$\neg g$	4, 2, 2	1, 1, 0

g

	g	$\neg g$
g	2, 2, 4	0, 1, 1
$\neg g$	1, 0, 1	1, 1, 1

$\neg g$

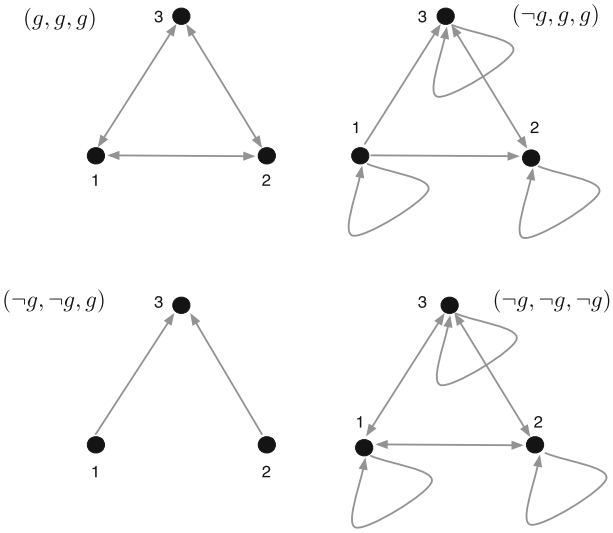

Fig. 4 Some BR-dependencies from the game matrix in Figure 3 (Example 5)

Example 5 (*Cycles in a three-person game*) Consider the following three-person game.[4] A committee of three jurors has to decide whether to declare a defendant in a trial guilty or not. All the three jurors want the defendant to be found guilty, however, all three prefer that the others declare the defendant while she declares her innocent. Also, they do not want to be the only ones declaring the defendant guilty if the other two vote for innocence. They all know each other's preferences. Figure 3 gives a payoff matrix for such game. Figure 4 depicts some cyclic BR-dependencies inherent in the game presented. Player 1 is *row*, player 2 *column*, and player 3 picks the right or left matrix. Among the ones depicted, (g, g, g) display s six cycles of length 3 and so does $(\neg g, \neg g, \neg g)$, which also contains three reflexive arcs—and hence, by Fact 1, is a Nash equilibrium. Also $(\neg g, g, g)$ is a Nash equilibrium, it does not contain any cycle of length 3, but it does contain two of length two between players 2 and 3. Finally, $(\neg g, \neg g, g)$ does not contain any cycle.

3.3 Reciprocity

We now proceed to isolate some specific forms of cycles. These will be used to define several variants of a property of strategic games which we call *reciprocity*. The idea is that, depending on the properties of the dependence cycles of a given profile, we can isolate some significant ways in which players are interconnected via a dependence relation. These will be linked, in the next section, to the existence of equilibria in appropriately transformed games.

[4] The game resembles a three-person variant of the prisoner's dilemma, where defection, although not being a dominant strategy, can turn out being a best response for all players, making defection—like in the prisoner's dilemma—a Nash equilibrium.

Definition 11 (*Reciprocity*) Let $\mathcal{G}$ be a game, σ a profile, and $\langle N, R^x_\sigma \rangle$ the corresponding dependence graph with $x \in \{BR, DS\}$. We say that:

(i) a profile σ is *x-reciprocal* if and only if there exists a partition $P(N)$ of N such that each element p of the partition is the orbit of some R^x_σ-cycle in $\langle N, R^x_\sigma \rangle$;
(ii) for $C \subseteq N$, a profile σ is *partially x-reciprocal in C* (or *C-x-reciprocal*) if and only if there exists a partition $P(C)$ of C such that each element p of the partition is the orbit of some R^x_σ-cycle in $\langle N, R^x_\sigma \rangle$;
(iii) a profile σ is *trivially x-reciprocal* if and only if $\langle N, R^x_\sigma \rangle$ is reflexive, that is, it contains $|N|x$-cycles whose orbits are singletons;
(iv) a profile σ is *fully x-reciprocal* if and only if $\langle N, R^x_\sigma \rangle$ contains at least one x-cycle with orbit N (i.e., a Hamiltonian cycle).

Let us explain the above definitions by considering the case of best response dependence (BR-dependence). A profile σ is BR-reciprocal if all players belong to some cycle of BR-dependence. This is the case in both the (U, L), i.e., 'cooperate-cooperate', and (D, R), i.e., 'defect-defect', outcomes in the prisoner's dilemma (see Figs. 1, 2). The other two outcomes are not BR-reciprocal as one of the two players does not belong to the orbit of any cycle.

Along the same lines, a profile σ is partially BR-reciprocal in coalition C (or C-BR-reciprocal) if all the members of C are partitioned by cycles of BR-dependence. This means, intuitively, that independently on whether the players outside of coalition C are linked by dependencies or not, the members of C are in a situation of reciprocity in which everybody plays a best response strategy for somebody else in the coalition. So, in the prisoner's dilemma, outcome (D, L)—maximally preferred by *row*—is $\{row\}$-BR-reciprocal as *row* is playing a best response for herself, hence being in a dependence relation with herself. A perfectly symmetric consideration can be made about (U, R) and *column*.

Finally, trivial and full BR-reciprocity are special cases of BR-reciprocity. In the first case all players belong to a reflexive arc, that is, all players play their own best response strategy. In the second case there exists one Hamiltonian cycle, that is, all players are connected to one another by a path of BR-dependence. For example, inspecting the BR-dependencies in the prisoner's dilemma (Fig. 2) it can be observed that: (U, L) is fully BR-reciprocal as it contains two Hamiltonian cycles: *row* $R^{BR}_{(U,L)}$ *column* $R^{BR}_{(U,L)}$ *row* and *column* $R^{BR}_{(U,L)}$ *row* $R^{BR}_{(U,L)}$ *column*. On the other hand, (D, R) is trivially BR-reciprocal as the only cycles are *column* $R^{BR}_{(D,R)}$ *column* and *row* $R^{BR}_{(D,R)}$ *row*.

To sum up, a profile is reciprocal when the corresponding dependence relation, be it a BR- or DS-dependence, clusters the players into non-overlapping groups whose members are all part of some cycle of dependencies (including degenerate ones such as reflexive links). It is partially reciprocal if its dependence graph contains at least one cycle. Trivial and full reciprocity refer to two extreme cases of reciprocity. In the first case the cycles are reflexive arcs and in the second case all players are 'visited' by one and the same cycle.

Before moving to the next section, we first provide one further illustrative example and then study the relation between the two types of reciprocity that arise from Definition 11: best response and dominant strategy reciprocity.

Example 6 (*Reciprocity in the three-person game*) Let us go back to Example 5 and to its BR-dependence graph given Fig. 4. The graph of profile (g, g, g) contains cycles which all yield the partition $\{\{1, 2, 3\}\}$ of the set of players. It is then a fully BR-reciprocal profile. The cycles of profile $(\neg g, g, g)$, instead, yield two partitions: $\{\{1\}, \{2\}, \{3\}\}$ and $\{\{1\}, \{2, 3\}\}$. So that profile is BR-reciprocal, but not fully BR-reciprocal. As its graph is reflexive, it is trivially BR-reciprocal, and also partially BR-reciprocal with respect each nonempty coalition.

Interestingly, profile $(\neg g, \neg g, \neg g)$ it is both fully and trivially BR-reciprocal. Intuitively, in that profile each player acts in favor of some other player by playing his/her own best response strategy. Finally, profile $(\neg g, \neg g, g)$ does not exhibit any form of reciprocity.

Here below we report a few relevant facts concerning the interplay between DS- and BR-reciprocity.

Fact 2 (DS- vs. BR-reciprocity) *Let $\mathcal{G}$ be a game, σ a profile, $C \subseteq N$, and $\langle N, R^x_\sigma \rangle$ be its dependence graph with $x \in \{BR, DS\}$. The following holds:*

(i) σ is C-BR-reciprocal iff σ_C is BR-reciprocal in $\mathcal{G} \downarrow \sigma_{\overline{C}}$;
(ii) σ is C-DS-reciprocal iff σ_C is DS-reciprocal in $\mathcal{G} \downarrow \sigma'_{\overline{C}}$ for any profile σ';
(iii) if σ is C-DS-reciprocal, then σ is C-BR-reciprocal, but not vice versa;
(iv) if σ is DS-reciprocal, then σ is BR-reciprocal, but not vice versa.

Proof [First claim] From left to right. By Definition 11, if σ is C-BR-reciprocal, then C is the union of orbits of R^{BR}_σ-cycles. That is, by Definition 9, each member i of C plays a best response to σ_{-i} for some member j in C. Notice that, if we consider player i, best responding to $\sigma_{C \setminus i}$ in $\mathcal{G} \downarrow \sigma_{\overline{C}}$ is equivalent to best responding to σ_{-i} in $\mathcal{G}$. From right to left. It follows directly by the notion of best response and by Definitions 11 and 8. [Second claim] From left to right. By Definition 11, if σ is C-DS-reciprocal, then C is the union of orbits of a R^{DS}_σ-cycle. That is, by Definition 9, each member i of C plays a dominant strategy for some member j in C. As a dominant strategy is such no matter what the other players do, the desired result follows directly. From right to left. It follows directly by the notion of dominant strategy and by Definitions 11 and 8. [Third claim] It follows from the fact that a dominant strategy is always a best response. [Fourth claim] It follows directly from the third claim by setting $C := N$. □

In words, the first claim states that a profile σ is partially BR-reciprocal in a coalition C if and only if the restriction σ_C of σ to C is BR-reciprocal with respect to the subgame (recall Definition 2) obtained from $\mathcal{G}$ by fixing the strategy of the complement $\overline{C}$ of coalition C. More concisely, a profile is partially BR-reciprocal in a given coalition if and only if it is BR-reciprocal in the subgame obtained by fixing what the players do who do not belong to the coalition. The second claim is similar and states that a profile is partially DS-reciprocal in a given coalition if and only if it is DS-reciprocal in all subgames obtainable by fixing the strategies of the players who are not in the coalition. These two claims show a first interesting difference between partial BR- and DS- reciprocity: partial BR-reciprocity is bound by the dependence structure of the current profile while partial DS-reciprocity is not. This is not surprising as the two forms of reciprocity build one on the notion of best response, and the other on the stronger notion of dominant strategy. A second important difference is pointed out by the third and fourth claims, which show that, as expected, that (partial) DS-reciprocity is a stronger notion than (partial) BR-reciprocity.

We now proceed to give a game theoretic interpretation of these definitions of reciprocity based on the notion of dependence cycle.

3.4 Reciprocity and equilibrium

We provide a characterization of reciprocity as defined in Definition 11 in terms of standard solution concepts. However, we first have to complement the set of notions provided in Sect. 2 with the notion of permuted game.

Definition 12 (*Permuted games*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game, σ a profile, and $\mu : N \mapsto N$ a bijection on N. The μ-permutation of game $\mathcal{G}$ is the game $\mathcal{G}^\mu = (N^\mu, S^\mu, \Sigma^\mu, \succeq_i^\mu, o^\mu)$ such that:

- $N^\mu = N$;
- $S^\mu = S$;
- for all $i \in N$, $\Sigma_i^\mu = \Sigma_{\mu(i)}$;
- for all $i \in N$, $\succeq_i^\mu = \succeq_i$;
- $o_\mu : \times_{i \in N} \Sigma_{\mu(i)} \to S$ is such that $o_\mu(\mu(\sigma)) = o(\sigma)$, where $\mu(\sigma)$ denotes the permutation of σ according to μ.

Intuitively, a permuted game $\mathcal{G}^\mu$ is therefore a game where the strategies of each player are redistributed according to μ in the sense that i's strategies become $\mu(i)$'s strategies, where players keep the same preferences over outcomes, and where the outcome function assigns same outcomes to same profiles.

Example 7 (*Two horsemen* [14]) "Two horsemen are on a forest path chatting about something. A passerby M, the mischief maker, comes along and having plenty of time and a desire for amusement, suggests that they race against each other to a tree a short distance away and he will give a prize of \$100. However, there is an interesting twist. He will give the \$100 to the owner of the slower horse. Let us call the two horsemen Bill and Joe. Joe's horse can go at 35 miles per hour, whereas Bill's horse can only go 30 miles per hour. Since Bill has the slower horse, he should get the \$100. The two horsemen start, but soon realize that there is a problem. Each one is trying to go slower than the other and it is obvious that the race is not going to finish. […] Thus they end up […] with both horses going at 0 miles per hour. […] However, along comes another passerby, let us call her S, the problem solver, and the situation is explained to her. She turns out to have a clever solution. She advises the two men to switch horses. Now each man has an incentive to go fast, because by making his competitor's horse go faster, he is helping his own horse to win!" [14, pp. 195–196].

Once the game of the example is depicted as the left-hand side game matrix in Fig. 5, it is possible to view the second passerby's solution as a bijection μ which changes the game to the right-hand side version. Now *row* can play *column*'s moves and *column* can play *row*'s moves. The result is a swap of (D, L) with (U, R), since (D, L) in $\mathcal{G}^\mu$ corresponds to (U, R) in $\mathcal{G}$ and vice versa. On the other hand, (U, L) and (D, R) stay the same, as the exchange of strategies do not affect them. As a consequence, profile (D, R), in which both horsemen engage in the race, becomes a dominant strategy equilibrium.

On the ground of these intuitions, it is possible to obtain a simple characterization of the different notions of reciprocity given in Definition 11 as the existence of equilibria in appropriately permuted games.

Theorem 1 (Reciprocity in equilibrium) *Let $\mathcal{G}$ be a game and $\langle N, R_\sigma^x \rangle$ be its dependence graph with $x \in \{BR, DS\}$ and σ be a profile. It holds that:*

(i) *σ is x-reciprocal iff there exists a bijection $\mu : N \mapsto N$ s.t. σ is a x-equilibrium in the permuted game $\mathcal{G}^\mu$;*

Fig. 5 The two horsemen game matrix and its permutation modeling the horse-swap

	L	R
U	0,0	1,0
D	0,1	1,0

$\mathcal{G}$

L	R
0,0	0,1
1,0	1,0

$\mathcal{G}^\mu$

(ii) – *σ is partially BR-reciprocal in C (or C-BR-reciprocal) iff there exists a bijection $\mu : C \mapsto C$ s.t. σ_C is a BR-equilibrium in the permuted subgame $(\mathcal{G} \downarrow \sigma_{\overline{C}})^{\mu}$;*
– *σ is partially DS-reciprocal in C (or C-DS-reciprocal) iff there exists a bijection $\mu : C \mapsto C$ s.t. σ_C is a DS-equilibrium in all permuted subgames $(\mathcal{G} \downarrow \sigma'_{\overline{C}})^{\mu}$ for any profile σ';*

(iii) *σ is trivially x-reciprocal iff σ is an x-equilibrium in $\mathcal{G}^{\mu}$ where μ is the identity over N;*

(iv) *σ is fully x-reciprocal iff there exists a bijection $\mu : N \mapsto N$ s.t.σ is a x-equilibrium in the permuted game $\mathcal{G}^{\mu}$ and μ is such that $\{(i, j) \mid i \in N \;\&\; j = \mu(i)\}$ is a Hamiltonian cycle in N.*

Proof The theorem states eight claims: four for $x = BR$ and four for $x = DS$. [First claim for $x = BR$.] From left to right, assume that σ is BR-reciprocal and prove the claim by constructing the desired μ. By Definition 11 it follows that there exists a partition P of N such that each element p of the partition is the orbit of some R_{σ}^{BR}-cycle. Given P, observe that any player i belongs to at most one member p of P. Now build μ so that $\mu(i)$ outputs the successor j (which is unique) of i in the cycle whose orbit is the p to which i belongs. Since each j has at most one predecessor in a cycle, μ is an injection and since domain and codomain coincide μ is also a surjection. Now it follows that for all i, j, $i R_{\sigma}^{BR} j$ implies, by Definition 9, that $\sigma_{\mu(i)}$ is a best response for i in σ. But in $\mathcal{G}^{\mu}$ it holds that $\sigma_{\mu(i)} \in \Sigma_i$ and since σ is reciprocal, by Definition 11, we have that for all $i\sigma_i$ is a best response in $\mathcal{G}^{\mu}$, and hence it is a Nash equilibrium. From right to left, assume μ to be the bijection at issue. It suffices to build the desired partition P from μ by an inverse construction of the one used in the left to right part of the claim. We set $i R_{\sigma}^{BR} j$ iff $\mu(i) = j$. The definition is sound w.r.t. Definition 9 because σ being a Nash equilibrium we have that $i R_{\sigma}^{BR} j$ iff j plays a best response for i in σ. Since μ is a bijection, it follows that R_{σ}^{BR} contains cycles whose orbits are disjoint and cover N. Therefore, by Definition 11, we can conclude that σ is BR-reciprocal. [First claim for $x = DS$.] The proof is analogous. [Second claim (i)] From right to left assume σ is C-x-reciprocal. It follows that there exists a non-empty C s.t. the restriction of R_{BR} to C is a cycle. By Definitions 2 and 11 this is equivalent to state that σ_C is BR-reciprocal in $\mathcal{G} \downarrow \sigma_{\overline{C}}$. From this, by the first claim we obtain that σ_C is a BR-equilibrium in $(\mathcal{G} \downarrow \sigma_{\overline{C}})^{\mu}$ for some bijection μ. As these steps are all equivalences (Fact 2 first equivalence) we directly obtain also the direction from left to right. [Second claim (ii)] The proof follows the line of the proof of the previous claim but exploits the second equivalence of Fact 2. [Third claim for $x = BR$] Follows directly from Definition 11 and Fact 1. [Third claim for $x = DS$] Follows directly from Definition 11 and Fact 1. [Fourth claim for $x = BR$] Follows directly from Definition 11 and the first claim. [Fourth claim for $x = DS$] It can be proven in the same way. □

Intuitively, the theorem connects all cycle-based forms of reciprocity identified in Definition 11 with equilibria in (sub-)games that could be obtained by appropriate permutations of the underlying game. Furthermore, the instructions for such permutations—which strategies go to which player—are provided by the existent cycles. So, if the profile is trivially reciprocal (third claim), then it is already an equilibrium, and if it is fully reciprocal (fourth claim), it then becomes an equilibrium via a permutation that follows one of the available Hamiltonian cycles over the set of players.

We hold Theorem 1 to be of particular interest for two reasons. First, it provides a clear connection between intuitions developed in the theory of dependence—such as the significance of cycles—with notions which lie at the hart of game theory—such as the one of

equilibrium. Second, it provides a systematic dependence-based rationale for modifications of games that allow desirable but unstable outcomes—such as the cooperative outcome in the prisoner's dilemma—to become equilibria.

3.4.1 Implementation via permutation

As just discussed, in view of Theorem 1, permutations can be fruitfully viewed as ways of *implementing*—in a social software sense [14]—reciprocal profiles. This terminology is worth casting in a definition.

Definition 13 (*Implementation as game permutation*) Let $\mathcal{G}$ be a game, σ a profile and $\langle N, R^x_\sigma \rangle$ be its dependence graph. Let also $\mu : N \mapsto N$ be a bijection with $C \subseteq N$. We say that:

(i) μ BR-implements σ iff σ is a BR-equilibrium in $\mathcal{G}^\mu$;
(ii) μ DS-implements σ iff σ is a DS-equilibrium in $\mathcal{G}^\mu$;
(iii) μ partially BR-implements σ in C iff σ_C is an BR-equilibrium in $(\mathcal{G} \downarrow \sigma_{\overline{C}})^\mu$;
(iv) μ partially DS-implements σ in C iff σ_C is a DS-equilibrium in $(\mathcal{G} \downarrow \sigma'_{\overline{C}})^\mu$ for any profile σ'.

Intuitively, implementation is here understood as a way of transforming a game in such a way that the desirable outcomes, in the transformed game, are brought about at an equilibrium point. In this sense we talk about BR- or DS-implementation. Analogously, partial BR- or DS-implementation consist in the realization of the desirable outcomes as equilibria in one subgame (partial BR-implementation) or all possible sub-games (partial DS-implementation).

Example 8 (*Implementation of cooperation in the prisoner's dilemma*) Just like for the two-horsemen example (Example 7) we can think of implementing the cooperative outcome (U, L) of the prisoner's dilemma (Fig. 1) by permuting the game according to the permutation μ dictated by one of the Hamiltonian cycles present in the dependence graph of that profile (Fig. 2): $\mu(row) = column$ and $\mu(column) = row$. In the resulting game where, essentially, *row* decides whether *column* plays L or R and *column* whether *row* plays U or D, the cooperative outcome is a Nash equilibrium by Theorem 1 (see Fig. 6). An example of partial BR-implementation is provided by Example 5 (see also Fig. 3). There, profile $(\neg g, g, g)$ is partially BR-reciprocal in coalition $\{1, 2\}$ (see Example 6). A permutation between 2 and 3 would yield a game such that (g, g) is a Nash equilibrium in the sub-game obtained by fixing the strategy of player 1 to $\neg g$ (notice, however, that in this case the identity permutation would also guarantee such result). In other words, was it so that 1 had already made his/her choice, swapping the strategies of 2 and 3 would lead to a stable outcome.

4 Solving dependencies: dependence games

The previous sections have shown how reciprocity can be given two corresponding formal characterizations: existence of cycles in a dependence structure, and existence of equilibria

Fig. 6 The prisoner's dilemma matrix and its permutation swapping *row* with *column* according to the cycle of profile (U, L)

$\mathcal{G}$

	L	R
U	2,2	0,3
D	3,0	1,1

$\mathcal{G}^\mu$

	L	R
U	2,2	3,0
D	0,3	1,1

in a suitably permuted game (Theorem 1). In the present section, we apply the notion of reciprocity to obtain a refinement of coalitional games. The intuition behind such refinement consists, in a nutshell, to allow coalitions to form only in presence of some sort of reciprocity. This will allow us to study a form of dependence-based cooperation which arises in strategic settings when the players' interests are somehow aligned.

A similar intuition can already be found in [19]:

> "As soon as there is a possibility of choosing with whom to establish parallel interests, this becomes a case of choosing an ally. When alliances are formed, it is to be expected that some kind of mutual understanding between the two players involved will be necessary. […] One can also state it this way: A parallelism of interests makes a cooperation desirable, and therefore will probably lead to an agreement between the players involved." [19, p. 221]

Once this intuitive notion of "parallelism of interests" is taken to mean "mutual dependence" [6] or "dependence cycle" [18] a bridge is laid between the theory of cooperative games and dependence theory, as a special case of cooperation. By pursuing this intuition in the light of the framework just presented, this section intends to show that dependence theory can feed some original insights into the theory of games.

4.1 Agreements

The central notion will be one of players' agreement. The key idea behind it is that, given a reciprocal profile (of some sort according to Definition 11), the players can fruitfully *agree* to transform the game by some suitable permutation of sets of strategies.

Definition 14 (*Agreements and partial agreements*) Let $\mathcal{G}$ be a game, (N, R^x_σ) be its dependence structure in σ with $x \in \{BR, DS\}$, and let $i, j \in N$. A pair (σ, μ) is:

(i) an x-agreement for $\mathcal{G}$ if σ is an x-reciprocal profile, and $\mu : N \mapsto N$ a bijection which x-implements σ;
(ii) a partial x-agreement in C (or a C-x-agreement) for $\mathcal{G}$, if σ is a C-x-reciprocal profile and $\mu : C \mapsto C$ a bijection which C-x-implements σ.

The set of x-agreements of a game $\mathcal{G}$ is denoted $x\text{-}AGR(\mathcal{G})$ and the set of partial x-agreements, that is the set of pairs (σ, μ) for which there exists a C such that μC-x-implements σ, is denoted $x\text{-}pAGR(\mathcal{G})$.

Intuitively, a (partial) agreement, of BR or DS type, can be seen as the result of coordination (endogenous, via the players themselves, or exogenous, via a third party like in Example 7) selecting a desirable outcome and realizing it by an appropriate exchange of strategies.

Example 9 (*Agreements in the prisoner's dilemma*) Let us go back to the prisoner's dilemma. Agreement $((D, R), \mu)$ with $\mu(i) = i$ for all players, is the standard DS-equilibrium of the strategic game. But there is another possible agreement, where the players swap their strategies: it is $((U, L), \nu)$, for which $\nu(i) = N \backslash \{i\}$. Here *row* plays cooperatively for *column* and *column* plays cooperatively for *row*. Of the same kind is the agreement arising in Example 7. Notice that in such example, the agreement is the result of coordination mediated by a third party (the second passerby). Analogous considerations can also be done about Example 5 where, for instance, $((g, g, g), \mu)$ with $\mu(1) = 2$, $\mu(2) = 3$, $\mu(3) = 1$ is a BR- agreement.

As we might expect, BR- and DS-agreements are related in the same way as BR- and DS-reciprocity (Fact 2). In what follows we will focus only on DS-agreements and partial DS-agreements so, whenever we talk about agreements and partial agreements, we mean DS-agreements and partial DS-agreements, unless stated otherwise.

4.2 Dominance

As there can be several possible agreements in a game, the natural issue arises of how to order them. We will do that by defining a natural notion of dominance between agreements, but first we need some auxiliary notions.

Definition 15 (*C-candidates and C-variants*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and C a non-empty subset of N. An agreement (σ, μ) for $\mathcal{G}$ is a *C-candidate* if C is the union of some members of the partition induced by μ, that is: $C = \bigcup X$ where X is an element of the partition induced by μ on N. An agreement (σ, μ) for $\mathcal{G}$ is a *C-variant* of an agreement (σ', μ') if $\sigma_C = \sigma'_C$ and $\mu_C = \mu'_C$, where μ_C and μ'_C are the restrictions of μ to C. As a convention we take the set of $\emptyset$-candidate agreements to be empty and an agreement (σ, ν) to be the only $\emptyset$-variant of itself.

In other words, an agreement (σ, μ) is a C-candidate if the partial dependence relation for σ of C and $\overline{C}$ follows exactly μ, and it is a C-variant of (σ', μ') if it differs from this latter at most in its C-part. We can now define the following notions of dominance between agreements and between partial agreements.

Definition 16 (*Dominance*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game and $C \subseteq N$ be a coalition. We say that:

(i) An *agreement* (σ, μ) is dominated iff for some coalition C there exists a C-candidate agreement (σ', μ') for $\mathcal{G}$ such that for all agreements (ρ, ν) which are $\overline{C}$-variants of (σ', μ'), $o(\rho) \succ_i o(\sigma)$ for all $i \in C$.

(ii) A *partial agreement* (σ_C, μ) in C is dominated iff for some coalition $D \subseteq N$ there exists (τ_D, ν) which is a D-agreement such that for all σ', τ', $o(\tau_D, \tau'_{\overline{D}}) \succ_i o(\sigma_C, \sigma'_{\overline{C}})$ for all $i \in D$.

The set of undominated agreements of $\mathcal{G}$ is denoted $DEP(\mathcal{G})$ and the set of undominated partial agreements is denoted $pDEP(\mathcal{G})$.

Intuitively, an agreement is undominated when a coalition C can force all possible agreements to yield outcomes which are better for all the members of the coalition, regardless of what the rest of the players can agree to do, that is, regardless of the $\overline{C}$-variants of their agreements. A partial agreement in coalition C is undominated when C can, by means of a partial permutation, force the game to end up in a set of states which are better for the member of the coalition no matter what the players in $\overline{C}$ do.

It is worth stressing the critical difference between the two notions of dominance. This difference resides in the fact that while dominance between agreements only considers deviations which are the results of agreements, dominance between partial agreements considers any form of possible deviation.

Example 10 (*Dominance between partial agreements*) In the three person game of Fig. 3, $((g_1, g_2), (\mu(1) := 2, \mu(2) := 1))$ is a partial DS-agreement in $\{1, 2\}$. This agreement, which represents a form of dependence-based cooperation between 1 and 2 dominates the partial DS-agreement in N—on a trivially DS-reciprocal profile—$((\neg g_1, \neg g_2, \neg g_3), (\mu(1) := 1,$

$\mu(2) := 2, \mu(3) := 3)$). In fact, it is undominated, since even the partial DS-agreement in $N((g_1, g_2, g_3), (\mu(1) := 2, \mu(2) := 3, \mu(3) := 1))$ (which is also a DS-agreement) does not dominate it.

4.3 Dependence-based coalitional games

Agreements exploit the dependence relations between the players in order to achieve some form of mutually beneficial cooperation.

It is then natural to use agreements to study games in strategic form as some form of coalitional games where players form coalitions only in presence of reciprocity. Standard questions of cooperative game theory can then be meaningfully asked, like the following one:

> *Can we characterize the notion of dominance for agreements and partial agreements (Definition 16) in terms of a suitable notion of stability in appropriately defined coalitional games?*

In order to answer this question we proceed as follows. First, starting from a game $\mathcal{G}$, we consider its representation $\mathcal{C}^{\mathcal{G}}$ as a coalitional game as illustrated in Sect. 2 (Definition 6). As Definition 6 abstracts from dependence-theoretic considerations we refine it in two ways, corresponding to the two different sorts of dependence upon which we want to build the coalitional game:

1. The first refinement is obtained by defining a coalitional game $\mathcal{C}^{\mathcal{G}}_{DEP}$ capturing the intuition that coalitions form only by means of *agreements* (Definition 14). Such games are called *dependence games*.
2. The second one is obtained by defining a coalitional game $\mathcal{C}^{\mathcal{G}}_{pDEP}$ capturing the intuition that coalitions form only by means of *partial agreements* (Definition 14). Such games are called *partial dependence games*.

Having done this, we show that the core of $\mathcal{C}^{\mathcal{G}}_{DEP}$ coincides with the set of undominated agreements of $\mathcal{G}$ (Theorem 2) and, respectively, that the core of $\mathcal{C}^{\mathcal{G}}_{pDEP}$ coincides with the set of undominated partial agreements of $\mathcal{G}$ (Theorem 3). We thereby obtain a game-theoretical characterization of Definition 16 formally linking agreements to the core of classes of coalitional games.

4.3.1 Dependence games

We start by refining the method to obtain a coalitional game from a game in strategic form (Definition 6), thus defining the notion of dependence game.

Definition 17 (*Dependence games from strategic ones*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. The dependence game $\mathcal{C}^{\mathcal{G}}_{DEP} = (N, S, E^{\mathcal{G}}_{DEP}, \succeq_i)$ of $\mathcal{G}$ is a coalitional game where the effectivity function $E^{\mathcal{G}}_{DEP}$ is defined as follows:

$$\begin{aligned} X \in E^{\mathcal{G}}_{DEP}(C) \Leftrightarrow\ & \exists \sigma_C, \mu_C \text{ S.T.} \\ & \exists \sigma_{\overline{C}}, \mu_{\overline{C}} : [((\sigma_C, \sigma_{\overline{C}}), (\mu_C, \mu_{\overline{C}})) \in AGR(\mathcal{G})] \\ & \text{AND } [\forall \sigma_{\overline{C}}, \mu_{\overline{C}} : [((\sigma_C, \sigma_{\overline{C}}), (\mu_C, \mu_{\overline{C}})) \in AGR(\mathcal{G}) \\ & \text{IMPLIES } o(\sigma_C, \sigma_{\overline{C}}) \in X]]. \end{aligned}$$

where $\mu : N \to N$ is a bijection.

This somewhat intricate formulation states nothing but that the effectivity function $E^{\mathcal{G}}_{DEP}(C)$ associates with each coalition C the states which are outcomes of agreements (and hence of reciprocal profiles), and which C can force via partial agreements (σ_C, μ_C) regardless of the partial agreements $(\sigma_{\overline{C}}, \mu_{\overline{C}})$ of $\overline{C}$.

We obtain the following theorem.

Theorem 2 (DEP vs. $CORE$) *Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. It holds that, for all agreements (σ, μ):*

$$(\sigma, \mu) \in DEP(\mathcal{G}) \;\Leftrightarrow\; o(\sigma) \in CORE(\mathcal{C}^{\mathcal{G}}_{DEP}).$$

where $\mu : N \to N$.

Proof [Left to right:] By contraposition, assume $o(\sigma) \notin CORE(\mathcal{C}^{\mathcal{G}}_{DEP})$. By Definition 7 this means that $\exists C \subseteq N, X \in E^{\mathcal{G}}_{DEP}(C)$ s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Applying Definition 17 we obtain that there exists an agreement $((\sigma'_C, \sigma'_{\overline{C}}), (\mu'_C, \mu'_{\overline{C}}))$ s.t. $\forall \sigma'_{\overline{C}}, \mu'_{\overline{C}}, o(\sigma'_C, \sigma'_{\overline{C}}) \in X$ and s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Now, $((\sigma'_C, \sigma'_{\overline{C}}), (\mu'_C, \mu'_{\overline{C}}))$ is obviously C-candidate, and all its C-variants yield better outcomes for C than σ. Hence, by Definition 16,$(\sigma, \mu) \notin DEP(\mathcal{G})$. [Right to left:] Notice that the set up of Definition 16 implies that, if (σ, μ) is dominated, then any other agreement for σ would also be dominated. So, by contraposition, assume $(\sigma, \mu) \notin DEP(\mathcal{G})$. By Definition 16, we obtain that there exists a C-candidate agreement (σ', μ') for $\mathcal{G}$ such that for all agreements (ρ, ν) which are $\overline{C}$-variants of (σ', μ'), $o(\rho, \nu) \succ_i o(\sigma)$ for all $i \in C$. But this means, by Definition 17, that $\exists C, X$ such that $X \in E^{\mathcal{G}}_{DEP}(C)$ and $x \succ_i o(\sigma)$ for all $x \in C$. Hence, by Definition 7, we obtain $o(\sigma) \notin CORE(\mathcal{C}^{\mathcal{G}}_{DEP})$. □

Put it otherwise, here is what Theorem 2 states. Given a game $\mathcal{G}$, a profile σ which is partially DS-implemented by μ (Definition 13) forms an undominated partial agreement (σ, μ) if and only if σ is in the core of the dependence game of $\mathcal{G}$. By taking Definition 11 and Theorem 1 into the picture, we thus see that Theorem 2 connects three apparently rather different properties of a strategic game $\mathcal{G}$: the existence of reciprocal profiles, the existence of DS-equilibria in permutations of $\mathcal{G}$, and the core of the dependence game built on $\mathcal{G}$.

4.3.2 Partial dependence games

We now move on to define the class of partial dependence game, in a way analogous to the one followed for dependence games in Definition 17.

Definition 18 (*Partial dependence games from strategic ones*) Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. The partial dependence game $\mathcal{C}^{\mathcal{G}}_{pDEP} = (N, S, E^{\mathcal{G}}_{pDEP}, \succeq_i)$ of $\mathcal{G}$ is a coalitional game where the effectivity function $E^{\mathcal{G}}_{pDEP}$ is defined as follows:

$$\begin{aligned} X \in E^{\mathcal{G}}_{pDEP}(C) \Leftrightarrow \exists \sigma_C, \mu_C \text{ S.T.} & \\ & (\sigma_C, \mu_C) \in pAGR(\mathcal{G}) \\ & \text{AND } [\forall \sigma_{\overline{C}} : o(\sigma_C, \sigma_{\overline{C}}) \in X]. \end{aligned}$$

where $\mu_C : C \to C$ is a bijection.

Partial dependence games are defined by just looking at the set of outcomes that each coalition can force by means of a partial agreement. Unlike Definition 17, Definition 18 is much closer to the standard definition of coalitional games based on strategic ones (Definition 6).

Like for dependence games, we have a characterization of the set of undominated partial agreements.

Theorem 3 ($pDEP$ vs. $CORE$) *Let $\mathcal{G} = (N, S, \Sigma_i, \succeq_i, o)$ be a game. It holds that, for all agreements (σ, μ):*

$$(\sigma, \mu) \in pDEP(\mathcal{G}) \;\Leftrightarrow\; o(\sigma) \in CORE(\mathcal{C}^{\mathcal{G}}_{pDEP}).$$

where $\mu : C \to C$ is a bijection with $C \subseteq N$.

Proof [Left to right:] By contraposition, assume $o(\sigma) \notin CORE(\mathcal{C}^{\mathcal{G}}_{pDEP})$. By Definition 7 this means that $\exists C \subseteq N, X \in E^{\mathcal{G}}_{DEP}(C)$ s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. Applying Definition 18 we obtain that there exists a partial agreement (ρ'_C, μ'_C) s.t. $\forall \rho'_{\overline{C}}, o(\rho'_C, \rho'_{\overline{C}}) \in X$ and s.t. $x \succ_i o(\sigma)$ for all $i \in C, x \in X$. By Definition 16, $(\sigma, \mu) \notin pDEP(\mathcal{G})$. [Right to left:] By contraposition, assume $(\sigma, \mu) \notin pDEP(\mathcal{G})$. By Definition 16, we obtain that there exists a partial agreement (ρ'_C, μ'_C) for $\mathcal{G}$ such that for all $\rho'_{\overline{C}}, o(\rho, \nu) \succ_i o(\sigma)$ for all $i \in C$. But this means, by Definition 18, that $\exists C, X$ such that $X \in E^{\mathcal{G}}_{pDEP}(C)$ and $x \succ_i o(\sigma)$ for all $x \in C$. Hence, by Definition 7, we obtain $o(\sigma) \notin CORE(\mathcal{C}^{\mathcal{G}}_{DEP})$. □

Like Theorem 2, Theorem 3 establishes a precise connection between the notions of partial reciprocity in a strategic game $\mathcal{G}$, the existence of DS-equilibria in all permuted subgames of $\mathcal{G}$, and the core of the partial dependence game built on $\mathcal{G}$.

4.4 Coalitional, dependence, partial dependence effectivity

The coalitional game $\mathcal{C}^{\mathcal{G}}$ built on a strategic game $\mathcal{G}$ and its dependence-based counterparts $\mathcal{C}^{\mathcal{G}}_{DEP}$ and $\mathcal{C}^{\mathcal{G}}_{pDEP}$ are logically related. The following fact shows how.

Fact 3 (Effectivity functions related) *The following relations hold:*

(i) For all $\mathcal{G}$: $E^{\mathcal{G}}_{pDEP} \subseteq E^{\mathcal{G}}$;
(ii) It does not hold that for all $\mathcal{G}$: $E^{\mathcal{G}}_{DEP} \subseteq E^{\mathcal{G}}_{pDEP}$; nor it holds that for all $\mathcal{G}$: $E^{\mathcal{G}}_{pDEP} \subseteq E^{\mathcal{G}}_{DEP}$;
(iii) It does not hold that for all $\mathcal{G}$: $E^{\mathcal{G}}_{DEP} \subseteq E^{\mathcal{G}}$; nor it holds that for all $\mathcal{G}$: $E^{\mathcal{G}} \subseteq E^{\mathcal{G}}_{DEP}$.

Proof [First claim] Suppose not. Then for some $\mathcal{G}$, some $X \subseteq S$ and some $C \subseteq N$ we have that $X \in E^{\mathcal{G}}_{pDEP}$ and $X \notin E^{\mathcal{G}}(C)$. The latter means that $\neg\exists \sigma_C \in \times_{i \in C} \Sigma_i$ such that $\forall \sigma_{\overline{C}} o(\sigma_C, \sigma_{\overline{C}}) \in X$. However this implies by elementary logical reasoning that $\neg\exists \sigma_C \in \times_{i\, inC} \Sigma_i \neg\exists \mu_C$ such that $(\sigma_C, \mu_C) \in pAGR(\mathcal{G})$ and $\forall \sigma_{\overline{C}} o(\sigma_C, \sigma_{\overline{C}}) \in X$, i.e. that $X \notin E^{\mathcal{G}}_{pDEP}$. Contradiction. [Second Claim] To refute the first inclusion consider a prisoner's dilemma in which $\{(D, R)\} \in E^{\mathcal{G}}_{DEP}(\{column, row\})$ but $\{(D, R)\} \notin E^{\mathcal{G}}_{pDEP}(\{column, row\})$. The second inclusion can be instead refuted by any game $\mathcal{G}$ that features a partial agreement (μ_C, σ_C) for players in C but no partial agreement for players in $\overline{C}$. [Third Claim] For the first inclusion consider a game $\mathcal{G}$ with tree players 1, 2, 3 and two actions $\{a, b\}$ for each of them. Suppose the only possible agreement is the identity permutation $\mu(i) = i$ and (a, a, a) is a DS-equilibrium. We have that $\{o(a, a, a)\} \in E^{\mathcal{G}}_{DEP}(\{1\})$ while $\{o(a, a, a)\} \notin E^{\mathcal{G}}(\{1\})$. For the second inclusion take $\mathcal{G}$ the prisoner's dilemma game in which $\{(U, R)\} \in E^{\mathcal{G}}(\{column, row\})$ but $\{(U, R)\} \notin E^{\mathcal{G}}_{DEP}(\{column, row\})$. □

The fact shows that dependence-based effectivity function considerably modify the powers assigned to coalitions by the standard definition of coalitional games on strategic ones

(Definition 6). Partial dependence effectivity functions instead really weaken the notion of coalitional ability, reducing the coalitional strategy at players' disposal. A formal consequence of Fact 3 is the establishment of the relation between $CORE(\mathcal{C}^{\mathcal{G}})$, $CORE(\mathcal{C}^{\mathcal{G}}_{DEP})$ and $CORE(\mathcal{C}^{\mathcal{G}}_{pDEP})$, as a direct consequence of the inclusion relation among their corresponding effectivity functions.

Summing up, the results in this section have shown that agreements and partial agreements are a form of coalitional power that can be related to standard cooperative solution concepts such as the core (Theorems 2, 3). In particular, partial agreements can be seen as a weakened forms of coalitional strategies (Fact 3), i.e. those strategies that can be executed only in presence of mutual reciprocity among the members of a coalition. As such partial dependence games, that generalize dependence games, should be understood as an intermediate level between the individualistic perspective studied in strategic games and the group perspective analyzed in coalitional games.

5 Further examples

5.1 An application to transferable utility games

The present section shows an application of dependence theory to transferable utility games [13]. In transferable utility games (shortly *TU games*) the preference relations are replaced by payoff functions, that associate to each strategy profile a positive real number, with the intuitive understanding that the number symbolizes what the player gets at the state associated to that strategy profile.

Definition 19 (*TU game*) A (strategic form) transferable utility game (TU game) is a tuple $\mathcal{G} = (N, \Sigma_i, p_i)$ where:

- N is a non-empty set of players;
- Σ_i is a set of strategies for player $i \in N$;
- $p_i : \prod_{i \in N} \Sigma_i \to \mathbb{R}_+$ is a payoff function, that associates to each player and each strategy profile a positive real number.

So, TU games are games in strategic form where the outcome function is substituted by a payoff function where numerical payoffs encode agents' preferences. All games in Fig. 1 are then examples of TU Games.[5]

Definition 6 allows us to translate strategic games into coalitional games by using so-called α-effectivity functions. In TU games this translation is not available as players together can only reach vectors of reals and not set of outcomes, but a similar notion can be defined, interpreting what a coalition can achieve as the best payoff that a coalition is able to achieve on its own, i.e., what we call the *value* of a coalition. We first define the coalitional payoff associated to each strategy profile.

[5] TU games are defined without an outcome function and the payoffs are directly associated to strategy profiles. TU games can be translated into standard strategic games (Definition 1) by endowing them with a bijective outcome function and a class preference relations $\preceq_i$ induced by the payoff functions as expected:

$$o(\sigma) \preceq_i o(\sigma') \text{ if and only if } p_i(\sigma) \leq p_i(\sigma')$$

for each strategy profile σ, σ' and each player $i \in N$.

Definition 20 (*Coalitional payoff in TU Games*) Let $\mathcal{G}$ be a TU game and $\sigma \in \Sigma$ be a strategy profile. The payoff of coalition C for the strategy profile σ, $p_C(\sigma)$ is defined as follows:

$$p_C(\sigma) = \sum_{i \in C} p_i(\sigma).$$

Taking into account the possible replies of the opponents, we are able to define the minimal payoff, namely $min_{\sigma_{\overline{C}}} p_C(\sigma)$, for each strategy σ_C that coalition C can play. In essence, we take a negative view on the opponents assuming that they will try to minimize C's payoff. Coalitions can then choose the best strategy knowing each minimal payoff, which constitutes the value of the coalition.

Definition 21 (*Value of a coalition*) Let $\mathcal{G}$ be a TU game and $C \subseteq N$ be a coalition. $v^{\alpha}(C)$, *the value* that coalition C is able to guarantee, is defined as follows:

$$v^{\alpha}(C) = max_{\sigma_C} min_{\sigma_{\overline{C}}} p_C(\sigma).$$

In words, the value of a coalition C is the payoff $p_C(\sigma)$ where σ_C is the most rewarding collective strategy that C can play knowing that $\sigma_{\overline{C}}$ is the toughest collective strategy by $\overline{C}$.

Definition 21 allows coalition C to select any collective strategy at its disposal. One immediate contribution of the theory of dependence is to restrict the set of available strategies via partial agreements. As done in Sect. 4, we restrict our attention to partial DS-agreements.

Definition 22 (*Coalitional Agreements in TU Games*) Let $\mathcal{G}$ be a TU game. σ_C is a *coalitional agreement of* C in game $\mathcal{G}$ if there exists $\mu : C \to C$, such that for all $i \in C$ and for all σ_i' and ρ_{-i} we have that:

$$p_{\mu(i)}(\sigma_i, \rho_{-i}) \geq p_{\mu(i)}(\sigma_i', \rho_{-i}).$$

We call $AGR^{TU}(\mathcal{G})_C$ the set of coalitional agreements of C at game $\mathcal{G}$.

Now we can define the negotiated value of a coalition, i.e., the payoff that a coalition can guarantee by undertaking agreements.

Definition 23 (*Negotiated value of a coalition*) Let $\mathcal{G}$ be a game and $C \subseteq N$ be a coalition. $v^{\alpha}_{DEP}(C)$, the *negotiated value* of coalition S is able to guarantee, is defined as follows:

$$v^{\alpha}_{DEP}(C) = max_{\sigma_C \in AGR^{TU}(\mathcal{G})_C} min_{\sigma_{\overline{C}}} p_C(\sigma).$$

Intuitively, $v^{\alpha}_{DEP}(C)$ represents the payoff that players in C can guarantee undertaking an agreement that answers the toughest collective strategy by their opponents. As a convention, if a coalition is not able to reach any agreement v^{α}_{DEP} is set to 0. Notice the similarity with partial agreements, where the opponents of a coalition can play strategies that *need not be themselves coalitional agreements*.

Several properties are desirable for TU games. One of the most fundamental one is that of superadditivity of the coalitional value, i.e the fact that coalitions can achieve more by uniting than by playing separately. In our case this translates into the following requirement: $v^{\alpha}_{DEP}(C) + v^{\alpha}_{DEP}(D) \leq v^{\alpha}_{DEP}(D \cup C)$ for each disjoint C, D. Under a mild assumption, this property holds.

Fact 4 *Let $C \subseteq N$, $D \subseteq N$ and $C \cap D = \emptyset$ and $v^{\alpha}_{DEP}(C) > 0$, $v^{\alpha}_{DEP}(C) > 0$.*

$$v^{\alpha}_{DEP}(C) + v^{\alpha}_{DEP}(D) \leq v^{\alpha}_{DEP}(D \cup C)$$

Proof It follows from the fact that $\sigma_C \in AGR^{TU}(\mathcal{G})_C$ and $\sigma_D \in AGR^{TU}(\mathcal{G})_D$ for disjoint C, D implies that $\sigma_{C \cup D} \in AGR^{TU}(\mathcal{G})_{C \cup D}$ and Definition 21. □

Fig. 7 Strangers on a train

	N	S	O
N	2,2	2,0	9,1
S	0,2	0,0	0,1
O	1,9	1,0	8,8

In words the proposition says that coalitions can favorably merge their partial agreements. In general partial agreements are not superadditive, as a consequence of the fact that bigger coalitions may hinder agreements instead of favoring them, unless its separate components could already agree on something. In the latter case a partial agreement can always be obtained, by merging the two disjoint partial agreements. Finally, it is instructive to notice that superadditivity would not be obtainable under simple agreements. This depends on the fact that agreements are defined with respect to a whole strategy profile limiting considerably the possibilities of coalition formation and, therefore, of coalition merging.

In summary, the concept of partial agreement applies naturally to games with transferable utility, where the value of coalitions can be calculated looking at the payoff that players can achieve together (Definition 21). In the context of such games, partial agreements turn out to be well behaved as they enforce, under a natural assumption, the desirable property of superadditivity (Fact 4) which constitutes a common assumption in cooperative game theory.

5.2 Strangers on a train

In this section we illustrate the building blocks of our theory of dependence within one all-encompassing example.

Example 11 (*Strangers on a Train*) In Patricia Highsmith's novel,[6] *Strangers on a Train* [11], that Alfred Hitchcock turned in 1951 into a movie with the same title, the following story takes place:

> Two protagonists wish to get out of an unhappy relationship. Architect Guy Haines wants to get rid of his unfaithful wife, Miriam, in order to marry the woman he loves, Anne Faulkner. Charles Anthony Bruno, a psychopathic playboy, deeply desires his father's death. On a train to see his wife, Guy meets Bruno, who proposes the idea of exchange murders: Bruno will kill Miriam if Guy kills Bruno's father; neither of them will have a motive, and the police will have no reason to suspect either of them.

We can illustrate our protagonists' setting, before any agreements are taken, with the two persons' matrix in Fig. 7.

In the example, both players have the same possibilities: either do nothing (N), commit the murder of their own significant other (S), or commit the murder of the other persons' significant other (O). Let Guy be the row player and Bruno the column player. Focusing on the choices of Guy, we notice that N is a *dominant strategy* for Guy (Definition 3), as whatever strategy Bruno plays, N is a best response to that strategy. For Bruno the reasoning pattern is symmetric, therefore his strategy N is also a dominant strategy. These two facts taken together mean that the strategy profile (N, N) is a *dominant strategy equilibrium* (Definition 4).

However the story takes an interesting twist once we consider what players could do for each other. The strategy O, by Guy, is a *dominant strategy for Bruno* (Definition 8), as it

[6] We thank Paul Harrenstein for having brought this example to our attention.

Fig. 8 Swapping murders

	N	S	O
N	2,2	0,2	1,9
S	2,0	0,0	1,0
O	9,1	0,1	8,8

is good for Bruno whatever Bruno himself decides to do. Same for Guy: the strategy O by Bruno is a *dominant strategy for Guy*. Once we identify what players can do for each other, the dependence relations can be automatically drawn: Guy DS*-depends* on Bruno for strategy O and on himself for strategy N (Definition 9), while Bruno DS*-depends on Guy* for strategy O and on himself for strategy N. Dependence *cycles* (Definition 10) suggests the possibility of reciprocal play: the profile (N, N), that is associated with two dependence cycles of length 1, is *trivially* DS*-reciprocal* (Definition 11), i.e. the only possible way for players to agree is to play for themselves, while the profile (O, O), that is associated with an hamiltonian dependence cycle, is *fully* DS*-reciprocal* (Definition 11), i.e. players can profit by playing for each other.

In this situation two *agreements* would be possible (Definition 14): $((N, N), \mu)$ and $((O, O), \nu)$, where μ is the identity permutation and ν is the players' transposition. However notice that the outcome resulting from (O, O) is preferred by both players to the outcome resulting from (N, N), which means that $((N, N), \mu)$ is a *dominated agreement* (Definition 16), while $((O, O), \nu)$ is undominated. The latter is also *stable* as, by Theorem 2, belongs to the core of the resulting *dependence game* (Definition 17). Therefore $((O, O), \nu)$ can be considered a *rational* outcome of the dependence game: Guy would find it reasonable to kill Bruno's father only if he knew that Bruno would kill his wife, and the same for Bruno. This would be possible if Guy could *lend* is action of killing *in exchange to* Bruno's one. The proposal of swapping murders, i.e. a simultaneous exchange of favors between the strangers, suggests itself. If this agreement could take place then the game would be transformed in the one pictured in Fig. 8, the transposition of the matrix in Fig. 7 under swap of strategies. The swap of players shown in this game DS-implements (O, O) (Definition 13).

We can finally observe that the game of the strangers on a train is also a TU game in the sense of Definition 19 and it also enjoys the property of superadditivity (Fact 4): the value of the coalition made by Guy and Bruno is strictly higher than the sum of the values made by Guy and Bruno disjoint.

6 Discussion

In this final section, we discuss related work, point at future directions of research, and conclude the paper.

6.1 Related work

We relate here our game-theoretic view of dependence to existing literature in MAS. To the best of our knowledge, almost no attention has been dedicated up till now to the relation between game theory and dependence theory. There are, however, three noteworthy exceptions:

– A study of the added value of exchanging tasks in a restricted game-like structure can be found in [2], where task-exchange is studied as a means to ease the computation of coalitions.
– Recently, the work presented in [4,16] elaborates on ideas close to [2] applying them to a special class of games, called Boolean games.

We first shortly describe the framework in [2] and then discuss [4,16].

6.1.1 Coalitions that exchange tasks

A series of papers by Boella, Sauro and van der Torre [1,2,15] put forward a formalization of the notions of power and dependence inspired by the work of Castelfranchi and colleagues. In their work the attempt of using dependence relations to form coalitional structures first makes its appearance. In [2] the authors frame their definitions within *task based power structures*.[7]

Definition 24 (*Task based power structures*)A task based power structure is a tuple $\langle Ag, G, T, goals, power\rangle$ where:

– Ag is a non-empty set of agents;
– G is a set non-empty of goals;
– T is a non-empty set of tasks;
– $goals : Ag \rightarrow 2^G$ is a function that associates to each agent in Ag the subset of goals G it desires to achieve;
– $power : 2^{Ag\times T} \rightarrow 2^G$ is a partial function that associates to a task assignment $\tau \subseteq Ag\times T$ a set of goals that the task assignment achieves.

From their very first definition, the structures analyzed in [2] are characterized by a clear resemblance to strategic games. However, a variety of new concepts are introduced, such as tasks and tasks assignments, power and goals. It is worth stressing, even more because these concepts are used also in [16], the primitive character of goals and tasks, which are introduced as sets with no further formal requirement. In a task based power structure the classical definitions of rationality (such as those in Definition 3) are not immediately available and the costs and the benefits for each agents need to be independently defined.

Definition 25 (*Costs and benefits*) The benefits and costs of the task assignment τ for agent $a \in Ag$ are defined as follows:

$$\text{benefits}(\tau, a) = goals(a) \cap power(\tau)$$
$$\text{costs}(\tau, a) = \{t \in T \mid (a, t) \in \tau\}$$

The authors of [2] proceed then defining a notion of domination among task assignments. A task assignment τ_1 is dominated if there is a task assignment $\tau_2 \subseteq \tau_1$ if all the agents involved in τ_2 enjoy higher benefits in τ_2 (or lower costs). Task assignments that are not dominated are called *do-ut-des* task assignments.

Despite the resemblance of task based power structures with strategic games, the do-ut-des links among players in [2] are of completely different nature from ours. Apart from the lack of structure of goals, the correspondent of preferences in tasks based structures, and the consequent difficulty of formulating classical solution concepts, it appears even more problematic

[7] Similar structures are studied in [1,15] and subsequent papers.

to relate the notion of benefit in Definition 25 to our notion of favor. While the latter is a straightforward generalization of the classical notions of best response and dominant strategy the do-ut-des in [2] is only concerned with the burden that a player bears in a task assignment and does not aim at being a formal correspondent of reciprocity in games. As authors of [2] rightly claim their definition should be interpreted as "give something to obtain something else", without concerning the strategic aspects of decision making.

6.1.2 Dependence in Boolean games

Recently, [4,16] have studied a notion of dependence for a restricted class of strategic games called Boolean games.

In [4], the authors study the sort of dependence relations between players which arise within a specific class of games called Boolean games [10]. In a nutshell, Boolean games are n-player games where players act by controlling the truth value of a propositional variable, and where players' preferences are dichotomous, that is, each player has a single goal—expressed by a propositional formula—which is either fulfilled or not. The work presented in [16] then extends some of the results presented in [4] to the class of cooperative Boolean games [8], that is, a coalitional version of Boolean games. Like in our case, the authors of [4,16] look at dependence relations as graph-theoretical information hidden within the game structure. However, there are several important differences.

First of all the simple structure of Boolean games, and in particular the fact that players' preferences are dichotomous, allows for a definition of dependence which is considerably simpler than ours (Definition 9):

> *Player i depends on player j if and only if j controls some propositional variables which are relevant for the satisfaction of i's goal.*

It is easy to see that such definition cannot be straightforwardly generalized to the case in which players have non-dichotomous preferences, as in that case it becomes unclear what the 'goal' of player i actually is. In fact, this is precisely the sort of issue that we went around by proposing Definition 9. Notice that, as a consequence, the two definitions of dependence differ radically in that the one proposed by [4,16] views dependence structures as properties of games, while ours view dependence structures as properties of the outcomes of games.

Secondly, it is worth mentioning an underlying difference in motivation between our work and the one presented in [4,16]. The latter develops the analysis of dependence relations essentially as a means to extract graphical information which eases the complexity of computing Nash equilibria in Boolean games and the core in cooperative Boolean games. What motivates our analysis instead, is rather the attempt to provide a game-theoretical foundation to dependence theory as such. This lead us to consider strategic games in their generality—rather than Boolean games—and to look at dependence as a means to characterize interesting properties of games (e.g., reciprocity) and to define a specific class of coalitional games, which has been the aim of Sect. 4.

6.2 Future work

The formalization of dependence relations and agreements provided here does not consider a variety of subtleties that might play a role in interaction. We list a few of them, sketching how to extend our framework in order to incorporate these more complex features.

6.2.1 Partial strategy permutation

Agreements are implemented by strategy permutation among stakeholders. If this operation fits perfectly games where players are endowed with a small number of strategies, such as those of Figs. 1 and 5, it seems more problematic when players are endowed with a larger number of strategies. Therefore, players may be interested in favors without necessarily having to lend control of all their actions. To this purpose, it would make sense to restrict possible permutations—exchanges of favors—to subsets of the available strategies. This could be done via a function that, when applied to a game $\mathcal{G}$, yields a game identical to $\mathcal{G}$, but where profiles are restricted to the available strategies, and where the outcome function is restricted accordingly. The intuition behind restricting the game is that players decide in advance the type of strategies that they allow to be agreed upon.

6.2.2 AND and OR dependence

Our definition of players' dependence allows for situations, as the one illustrated in Fig. 3, where a player can be simultaneously dependent on several other players, suggesting the possibility of many possible agreements. In the literature on dependence theory (cf. [18]) this form of dependence is usually referred to as *OR dependence*, as opposed to *AND dependence*, where instead a player is dependent on the *combined strategy* of other players, i.e. a sort of dependence not on a player but on a coalition. While the first can be easily accommodated in our framework, for the latter a generalization is required, that allows a dependence relation between a player and a coalition. The informal account in [18] suggests that AND dependencies and OR dependencies have different consequences for the stability of coalitions. If a situation of AND dependence of player i on players j and k grants the latter two players a power position (as i needs both), a situation of OR dependence allows player i to choose among the possible stakeholders in a possible agreement: in some sense players profit from OR dependencies. A suitable generalization of the definition of dependence should be able to account for this feature.

6.2.3 Extensive interaction

Dependence and agreements have been formulated for strategic games, where decisions have one-shot nature and no temporality is involved. However dependencies are naturally present in extensive interaction as well and agreements make perfect sense there. In order to analyze dependence in extensive games we can always adopt the standard translation of an extensive game into a strategic one [13]. Dependence relations and agreements can then be retrieved in the usual way, by resorting to the strategic game we have obtained. However extensive games have special features. Their typical solution concept, for instance, is that of *sub-game perfect equilibrium*, i.e. a Nash-equilibrium that rules out incredible threats [13]. What is interesting for a theory of dependence in extensive interaction is whether analogue solution concepts can be obtained for dependence relations. A generalization of the notion of sub-game perfect equilibrium to a notion of *sub-game perfect equilibrium for someone else* could be obtained as a refinement of the notion of equilibrium based on best response for someone else that we have studied here. Such refinement should take care of ruling out strategies determining incredible favors.

6.3 Conclusion

The contribution of the present work is two-fold:

- First and foremost it has been shown that the intuitive notion of dependence relation originating from social and cognitive science literature [5,7] can be fully incorporated within the theory of games, contributing to the construction of solution concepts that account for its underlying dynamics. The standard solution concepts of best response and dominant strategies first provided in Definition 3 have been generalized to best response and dominant strategy for someone else in Definition 8, providing a basis for formalizing reciprocity in games.
- Second, once the gametheoretical account has been laid, it has been shown that central dependence-theoretic notions such as the notion of cycle have natural game-theoretic correspondents (Theorem 1). Furthermore, dependence theory has been demonstrated to give rise to types of cooperative games where solution concepts such as the core can be applied. The relation between the various forms of cooperative games where coalitions undertake agreements (dependence and partial dependence) have been analyzed, together with the dominance they induce on agreements (Theorems 2, 3). The results suggest the presence a full spectrum of cooperative solution concepts for dependence structure, that form a partial order under the inclusion relation, whose further investigation pose an interesting research challenge.

All in all, the paper has shown that informal notions elaborated in the social sciences, such as dependence, can profitably be incorporated within a formal game-theoretical framework, contributing to refine the analysis of strategic and coalitional games.

Acknowledgments The authors would like to thank the anonymous reviewers of AAMAS'10 and COMSOC'10, where previous versions of the work have been presented, and the anonymous reviewers of JAAMAS. Their remarks have been of great help for developing the present version of the work. Davide Grossi wishes to acknowledge support by the *Netherlands Organisation for Scientific Research* (NWO) under the VENI grant 639.021.816.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Boella, G., Sauro, L., & van der Torre, L. (2005). Admissible agreements among goal-directed agents. In *Proceedings of 2005 IEEE/WIC/ACM international conference on intelligent agent technology (IAT'05)* (pp. 543–554). USA: IEEE Computer Society.
2. Boella, G., Sauro, L., & van der Torre, L. W. N. (2005). Reducing coalition structures via agreement specification. In F. Dignum, V. Dignum, S. Koenig, S. Kraus, M. P. Singh, & M. Wooldridge (Eds.) *AAMAS* (pp. 1187–1188). New York: ACM.
3. Boella, G., Sauro, L., & van der Torre, L. (2006) Strengthening admissible coalitions. In *Proceeding of the 2006 conference on ECAI 2006: 17th European conference on artificial intelligence* (pp. 195–199). New York: ACM.
4. Bonzon, E., Lagasquie-Schiex, M.-C., & Lang, J. (2009). Dependencies between players in boolean games. *International Journal of Approximate Reasoning, 50*, 899–914.
5. Castelfranchi, C. (1998). Modelling social action for AI agents. *Artificial Intelligence, 103*, 157–182.
6. Castelfranchi, C., Cesta, A., & Miceli, M. (1992). Dependence relations among autonomous agents. In E. Werner & Y. Demazeau (Eds.), *Decentralized A.I.3*. Amsterdam: Elsevier.
7. Coleman, J. (1990). *Foundations of social theory*. Cambridge: Belknap Harvard.

8. Dunne, P., van der Hoek, W., Kraus, S., & Wooldridge, M. (2008). Cooperative boolean games. In *Proceedings of AAMAS 2008* (pp. 1015–1022). New York: ACM.
9. Grossi, D., & Turrini, P. (2010) Dependence theory via game theory. In W. van der Hoek & G. Kaminka (Eds.), *Proceedings of AAMAS 2010*. New York: ACM.
10. Harrenstein, P., van der Hoek, W., Meyer, J.-J.Ch., & Witteveen, C. (2001). Boolean games. In J. van Benthem (Ed.), *Proceedings of TARK'01* (pp. 287–298). San Fransisco: Morgan Kaufmann.
11. Highsmith, P. (1950). *Strangers on a train*. USA: W.W. Norton.
12. Moulin, H., & Peleg, B. (1982). Cores of effectivity functions and implementation theory. *Journal of Mathematical Economics, 10*, 115–145.
13. Osborne, M. J., & Rubinstein, A. (1994). *A course in game theory*. Cambridge: MIT Press.
14. Parikh, R. (2002). Social software. *Synthese, 132*(3), 187–211.
15. Sauro, L. (2006). Qualitative criteria of admissibility for enforced agreements. *Computational & Mathematical Organization Theory, 12*(2-3), 147–168.
16. Sauro, L., van der Torre, L., & Villata S. (2009) Dependency in cooperative boolean games. In A. Håkansson, N. Nguyen, R. Hartung, R. Howlett, & L. Jain (Eds.), *Proceedings of KES-AMSTA 2009*. *LNAI* (Vol. 5559, pp. 1–10). Heidelberg: Springer
17. Sichman, J. (1998) Depint: Dependence-based coalition formation in an open multi-agent scenario. *Journal of Artificial Societies and Social Simulation, 1*(2).
18. Sichman, J., & Conte, R. (2002). Multi-agent dependence by dependence graphs. In *Proceedings of AAMAS 2002* (pp. 483–490). New York: ACM.
19. von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.