Strategic Games with Goal-Oriented Strategies*

Marek Hudík†

IBSS, Xi'an Jiaotong-Liverpool University, Suzhou

Center for Theoretical Study, Charles University, Prague

Abstract

This article models player strategies as goal-oriented and introduces a new solution concept characterized by mutual compatibility of players' goal-oriented strategies – Hayek equilibrium. Hayek equilibrium is understood as a complementary solution concept to Nash equilibrium: If an outcome is a Nash equilibrium but not Hayek equilibrium, then this outcome may be unstable "from without", as the players may have an incentive to change the game. On the other hand, if an outcome is a Hayek equilibrium but not a Nash equilibrium, then the outcome is appealing to players; yet, it is unstable within the game, as the players can profitably deviate from this outcome. Several applications of the model with goal-oriented strategies are discussed: It is shown that the concept of Hayek equilibrium can help to explain cooperation in the Prisoner's Dilemma. Furthermore, an explicit modeling of players' goals allows for more adequate definition of "pure conflict", "pure common-interest" and "mixed-motive" games. Finally, it is argued that goal-orientedness can be considered as one of the unifying concepts of behavioral sciences.

Keywords: goal-oriented strategies, Hayek equilibrium, pure conflict games, pure common-interest games, mixed-motive games, cooperation in the Prisoner's Dilemma, unification of behavioral sciences

JEL: C70, C72

---

* Earlier versions of this paper were presented at the Prague Conference on Political Economy (4/2014) and at the research seminars at the Center for Theoretical Studies (6/2014) and at the University of Nottingham Ningbo China (5/2015). I am grateful to the participants at these events for their helpful comments, especially to David Storch, Shravan Luckraz, Sailesh Gunessee, and Lu Dong. I also greatly benefited from the comments by Petr Bartoň, Antonín Machač, Jirka Lahvička and Petr Špecián. Any mistakes are, of course, mine alone.

† Marek Hudík; Xi'an Jiaotong-Liverpool University, 111 Ren Ai Road, Suzhou, China; Center for Theoretical Study, Charles University, Jilská 1, Prague, 110 00, Czech Republic. E-mail: marek.hudik@xjtlu.edu.cn.

Strategic Games with Goal-Oriented Strategies

Payoffs in strategic games may be interpreted as conveying two pieces of information: firstly, they reflect how successful a strategy is in reaching the goals that the player has in mind, and secondly, they reflect how valuable these goals are to the player. In many applications of game theory, uncovering these two pieces of information is not necessary; however, as argued in this paper, there are situations in which the question of what is behind the players' payoffs is of interest. What are then the benefits of explicit modeling of players' goals?

First of all, it allows for analyzing decision procedures of players who make choices with certain goals in mind. Players' goal-oriented plans of action may or may not be mutually compatible: for instance, a soccer player's strategy to "kick to the left to score a goal" is not compatible with the goalie's strategy to "jump to the left to prevent a goal". On the other hand, seller's strategy to "sell *X* for at least \$1", is compatible with a buyer's strategy to "buy *X* for at most \$2". Intuitively, players' strategies are mutually compatible, if both players are successful in achieving their goals. An advantage of this approach is that players do not have to be aware what game they are playing: They only observe whether they were successful or not without knowing choices of the others or the state of nature.

To account for this compatibility of player's goal-oriented strategies, I introduce a solution concept which I call "Hayek equilibrium" as Hayek (1937) was the first to introduce (in a different context) the notion of equilibrium as "compatibility of plans". Hayek equilibrium represents a notion complementary to that of Nash equilibrium: In particular, I argue that if an outcome is a Nash equilibrium but not a Hayek equilibrium, then this outcome tends to be unstable "from without", as players who are not successful in achieving their goals may have an incentive to change the game. The idea of instability from without is especially relevant for decision making in real-world setting where rules of the game are usually not simply "given" but can be modified by players' actions. If, on the other hand, an outcome is a Hayek equilibrium but not a Nash equilibrium, then it is appealing to players because they are successful in achieving their goals; yet, it is unstable within the game, as the players can profitably deviate from this outcome (i.e. attain a more valuable goal). Thus, Hayek equilibrium can, for instance, help to explain why many people cooperate in one-shot Prisoner's Dilemma in laboratory experiments (Colman 1995) as well as when playing outside laboratory (List 2006). This explanation differs from framing-based explanations proposed elsewhere (Bacharach 2006; Sugden 2000; Bicchieri 2005). Most importantly, my explanation does not require reconsideration of the Nash equilibrium concept.

Another advantage of explicit modeling of players' goals is that it allows for more satisfactory definition of "pure conflict game", "pure common-interest game" and "mixed-motive game". I give examples showing that the traditional classification based solely on relationship between players' payoffs (Schelling 1980) is inadequate. For instance, a game which appears to be of a pure conflict under the traditional definition, may actually involve elements of common interest if players' goals are examined. I show that this fact has also practical implications: a conflict in which no players' goals are mutually compatible may require different policy measures than a situation in which some of players' goals are mutually compatible.

Last but not least, I argue that goal-orientedness can be considered as one of the unifying concepts of behavioral sciences – in fact, it is already employed in many disciplines, including psychology (Locke and Latham 2002; 2013), biology (Mayr 1988; 1992), and cybernetics and systems theory (Rosenblueth et al. 1943; Ashby 1957; Bertalanffy 1968). Moreover, as shown below, the idea of goal-orientedness is also consistent with maximizing behavior traditionally used in economics. The model introduced in this article can thus provide a link between the rational choice theory and other approaches to behavior and hence contribute to the attempts to unify all behavioral sciences (most notably Gintis 2009).

Game theoretic literature on modeling players' goals is small.[1] Although various authors do sometimes speak about goals,[2] formal models are usually lacking. One exception proving the rule is Castelfranchi and Conte (1998) who explore the issue of applicability of game theory to artificial intelligence problems and propose what they call "goal-based strategy" as an alternative to utility maximization; unfortunately, they do not develop the idea any further. Apart from this proposal, they also correctly observe that strategies are sometimes (implicitly or explicitly) described as goal-oriented: thus for instance one of the strategies in the Prisoner's Dilemma is usually described as "cooperate" indicating that the outcome aimed at is cooperation.[3] My model is consistent with Castelfranchi and Conte's (1998) proposal but contrary to these authors, I argue that the concept of goal-oriented strategies is compatible with utility maximization and can be incorporated into the standard game-theoretic models.

[1] This is however not true for economics literature in general: probably the best-known model of purposeful behavior is Becker's (1998) model of consumption as production of commodities. For a survey of this literature, see e.g. Dietrich and List (2013b).

[2] For instance, the concept of forward induction of Kohlberg and Mertens (1986) is based on goal-based reasoning.

[3] Another standard example is the Stag Hunt game, where the strategies are typically described with goals that players' want to achieve (i.e. "Stag" and "Hare").

Somewhat related ideas to ones explored in this article are found in Dietrich and List (2013b; 2013a). The authors construct a model of “reason-based rational choice” in which players’ payoffs are derived from their “motivational states”. If the motivational state changes, then the player’s payoffs may change as well (see also Hudík (2014) for a discussion of this model). The model introduced in this article is in this respect similar; in particular, “motivational state” is interpreted as a set of goals. This specification has an intuitive appeal and allows an analysis of mutual compatibility of strategies with interesting applications.

In the next section I introduce the concept of goal-oriented strategy and I formally define strategic games with goal-oriented strategies.

1 Model

Let N be a finite set of players. For each $i \in N$, define a non-empty set of actions A_i, and a non-empty set of goals G_i. To capture the notion of goal-orientedness of behavior, define for each player $i \in N$ a set of goal-oriented strategies $S_i \subseteq A_i \times (2^{G_i} \setminus \emptyset)$. This definition allows for the possibility that an action is associated with multiple goals as well as for the possibility that different actions are associated with the same goal. The set of strategy profiles $\times_{j \in N} S_j$ is denoted by S.

Whether player achieves her goals or not depends not only on the strategies taken by her and other players but also on the state of nature. To account for this, define a finite set of states Ω

and a probability measure *p* on Ω. It is assumed that in every state $\omega \in \Omega$, the signal that a player observes is always the same.[4]

To assess to what extent are goals of a player compatible with goals of other players in a given state, define for each $i \in N$ a success function $c_i : S \times \Omega \to \{0,1\}^{|G_i|}$ which assigns to each strategy profile in every state of nature a $|G_i|$-tuple of probabilities $c_i(g_i \mid s, \omega)$ specifying for each goal $g_i \in G_i$ whether the player *i* achieves her goals (probability 1) or not (probability 0), if the outcome is (s, ω). For each strategy profile *s*, the success function c_i, together with the probability measure *p* over the states, generates a bundle $\lambda_i(s)$ which assigns to each $g_i \in G_i$ an overall probability $\lambda_i(g_i \mid s)$ that g_i is achieved by *i* given the strategy profile *s*. This probability is referred to as probability of success of g_i and is calculated as $\lambda_i(g_i \mid s) = \sum_{\omega \in \Omega} p(\omega) c_i(g_i \mid s, \omega)$. For each player *i*, denote the set of the probability bundles $\lambda_i(s)$ by L_i.

Finally, since each goal may have a different importance to a player, define for each player $i \in N$ a complete and transitive preference relation $\succsim_i$ on the set L_i.[5] As usual, preferences can be conveniently represented by a payoff function defined in the standard way.

[4] It is, of course, possible to drop this restriction and to construct a more general model. Nevertheless, the aim of this article is to introduce a model that can be used to represent the same situations typically modeled as strategic games.

[5] Note that it is assumed that players care only about the overall probability of success of their goals. In particular, they do not distinguish between decrease in probability of success due to choices of the other players and due to chance. These two cases can be treated separately by defining preferences over the set of probability measures over $S \times \Omega$.

Game with goal-oriented strategies is then defined as follows.

DEFINITION 1: *A **strategic game with goal-oriented strategies** is an octuple* $\langle N, (A_i), (G_i), (S_i), \Omega, p, (c_i), \succsim_i \rangle$

Definition 1 can be illustrated with the following version of the Hawk-Dove game.

EXAMPLE 1: *Let* $N = \{1, 2\}$, $A_1 = A_2 = \{Hawk, Dove\}$,

$G_1 = G_2 = \{Get\ the\ prey,\ Avoid\ conflict\}$, *and*

$S_1 = S_2 = \{(Hawk;\ Get\ the\ prey,\ Avoid\ conflict), (Dove;\ Get\ the\ prey,\ Avoid\ conflict)\}$.

There are four states of nature, each occurring with the probability 1/4:

$\Omega = \{11, 12, 21, 22\}$. *For instance,* 12 *denotes that the player 1 gets the prey if both play Hawk and the player 2 gets the prey if both play Dove. Success functions for each state are shown in Fig. 1(a)-(d), and probabilities of success and payoffs are shown in the Fig. 1(e) and 1(f) respectively.*

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	(1, 0), (0, 0)	(1, 1), (0, 1)
$(D;G,A)$	(0, 1), (1, 1)	(1, 1), (0, 1)

(a) $\omega = 11$

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	(1, 0), (0, 0)	(1, 1), (0, 1)
$(D;G,A)$	(0, 1), (1, 1)	(0, 1), (1, 1)

(b) $\omega = 12$

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	(0, 0), (1, 0)	(1, 1), (0, 1)
$(D;G,A)$	(0, 1), (1, 1)	(1, 1), (0, 1)

(c) $\omega = 21$

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	(0, 0), (1, 0)	(1, 1), (0, 1)
$(D;G,A)$	(0, 1), (1, 1)	(0, 1), (1, 1)

(d) $\omega = 22$

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	(1/2, 0), (1/2, 0)	(1, 1), (0, 1)
$(D;G,A)$	(0, 1), (1, 1)	(1/2, 1), (1/2, 1)

(e) *Probabilities of success*

	$(H;G,A)$	$(D;G,A)$
$(H;G,A)$	0, 0	3, 1
$(D;G,A)$	1, 3	2, 2

(f) *Payoffs*

Fig. 1: *Hawk-Dove game with goal-oriented strategies*

Solution concepts for strategic games with goal-oriented strategies are defined in the following section.

2 Solution concepts

I define two solution concepts for strategic games with goal-oriented strategies: Nash equilibrium and Hayek equilibrium.

DEFINITION 2: *A **Nash equilibrium of a strategic game with goal-oriented strategies** $\langle N,(A_i),(G_i),(S_i),\Omega,p,(c_i),\succsim_i\rangle$ is a profile $s^* \in S$ of goal-oriented strategies with the property that for every player $i \in N$ we have $\lambda_i(s_i^*, s_{-i}^*) \succsim_i \lambda_i(s_i, s_{-i}^*)$ for all $s_i \in S_i$.*

The definition of Nash equilibrium is standard; the only difference between the conventional strategic games and strategic games with goal-oriented strategies is that the latter approach specifies what is "behind" payoffs. However, uncovering this information allows to introduce an additional solution concept, based on the considerations whether players are successful in attaining their goals.

DEFINITION 3: *A goal-oriented strategy $s_j' \in S_j$ is **perfectly successful** in the outcome (s_j', s_{-j}) if $\lambda_j(g_j \mid s') = 1$ for every g_i associated with s_j'.*

DEFINITION 4: *A **Hayek equilibrium of a strategic game with goal-oriented strategies** $\langle N,(A_i),(G_i),(S_i),\Omega,p,(c_i),\succsim_i\rangle$ is a profile $\hat{s} \in S$ of goal-oriented strategies with the property that for each $i \in N$, $\hat{s}_i$ is perfectly successful in $\hat{s}$.*

Nash equilibrium represents a stable outcome *within* a game: No player can profitably deviate from this outcome. Nevertheless, I claim that if Nash equilibrium is not at the same time a Hayek equilibrium, then this Nash equilibrium may be unstable "from without": players whose equilibrium strategies are not perfectly successful may have an incentive to change the game, e.g. by taking strategic moves or by searching for new strategies to achieve their goals.

To show this, consider the version of Hawk-Dove game of the Example 1. Nash equilibria are ((*H;G,A*), (*D;G,A*) and ((*H;G,A*), (*D;G,A*)). However, in both Nash equilibria only the strategy (*H;G,A*) of one of the players is perfectly successful. Therefore, the player whose equilibrium strategy is (*D;G,A*) has an incentive to modify the game. Admittedly, one can argue that this incentive can be inferred also from the players' payoffs (Fig. 1(f)): the player whose equilibrium payoff is 1 may aim at changing the game in some way in order to obtain either 2 or 3. The advantage of the approach proposed in this article, is that the instability from without can be inferred directly from the profile $\lambda_i(s)$, i.e. without comparing it with other profiles in the game. This is relevant in particular when my approach is extended to situations in which players' sets of strategies are not fixed and players can search for new and potentially more successful strategies.

Is it possible that no strategy in a Nash equilibrium is perfectly successful? And on the contrary, is it possible to have a Hayek equilibrium that is not a Nash equilibrium? The answer to both questions is in affirmative as illustrated by the next example.

EXAMPLE 2: *Let* $N = \{1, 2\}$, $A_1 = A_2 = \{C, D\}$, $G_1 = G_2 = \{\$3,\ \$2,\ \$1\}$, $S_1 = S_2 = \{(C,\$2), (D,\$3)\}$ *and let there be only one state of nature. Probabilities of success and payoffs are shown in the Fig. 2(a) and 2(b) respectively.*

	(*C*,\$2)	(*D*,\$3)
(*C*,\$2)	(0, 1, 0), (0, 1, 0)	(0, 0, 0), (1, 0, 0)
(*D*,\$3)	(1, 0, 0), (0, 0, 0)	(0, 0, 1), (0, 0, 1)

(a) Probabilities of success

	(*C*,\$2)	(*D*,\$3)
(*C*,\$2)	2, 2	0, 3
(*D*,\$3)	3, 0	1, 1

(b) Payoffs

Fig. 2: *Prisoner's Dilemma with goal-oriented strategies*

The game in the Example 2 is a Prisoner's Dilemma with Nash equilibrium ((*D*,\$3), (*D*,\$3)). In this Nash equilibrium neither strategy is perfectly successful and both players have an incentive to change the game. On the other hand, the outcome ((*C*,\$2), (*C*,\$2)) is a Hayek equilibrium. Although this outcome is not stable within the game as it is not a Nash equilibrium, it is appealing to the players because they are successful in attaining the goal they have in mind. As mentioned, it has been observed that many people actually choose to cooperate in a Prisoner's Dilemma. The notion of Hayek equilibrium may be one of the explanations of the observed play.[6] Nevertheless, to what extent this is the case, cannot be assessed without empirical tests.

The following section shows some additional applications of explicit modelling of players' goals. Firstly, I demonstrate that the concept of goal-oriented strategy allows for a more adequate definition of "pure conflict", "pure common-interest" and "mixed-motive" games. Subsequently, I argue that the notion of goal-orientedness can provide a link between economics and other disciplines, in particular biology.

3 Applications

3.1 Classification of games[7]

Schelling (1980) introduced a classification of games which distinguishes among "pure conflict" (or "zero-sum"), " pure common-interest" (or "pure-coordination") and "mixed-

[6] Analogously, the concept of goal-based reasoning and the concept of Hayek equilibrium may explain some of the non-Nash-equilibrium play in other games, such as the traveler's dilemma game (Basu 1994; Goeree and Holt 2001), or various versions of the centipede game (Rosenthal 1981; McKelvey and Palfrey 1992; Beard and Beil 1994; Palacios-Huerta and Volij 2009).

[7] This section is based on Hudik (2015).

motive" games. The definition of these categories is based on relationships between payoffs of various players: If players' payoffs are perfectly positively correlated, then the game is of pure common-interest (Fig. 1(a)); if the payoffs are perfectly negatively correlated, then the game is of pure conflict game (Fig. 1(b)). Mixed-motive games are those in which players' payoffs are imperfectly correlated.

	X	*Y*
X	1, 1	0, 0
Y	0, 0	1, 1

(a)

	X	*Y*
X	1, 0	0, 1
Y	0, 1	1, 0

(b)

Fig. 3: *Pure common-interest and pure conflict game*

Although this payoff-based definition seems plausible, it is sometimes inadequate as shown by the following example:

EXAMPLE 3 (*A dating game with mixed motives*): *Consider two players, John and Blonde. John wants to meet with Blonde in a bar but he also wants to meet with another person, Brunette. Blonde wants to meet with John but she also wants to prevent John from meeting with Brunette. Both John and Blonde choose between two bars, X and Y. Blonde and Brunette are never in the same bar and so John always meets with one or the other. If John prefers meeting with Blonde to meeting with Brunette, then the game is a pure common-interest game such as the one represented in Fig. 3(a); If John prefers meeting with Brunette to meeting with Blonde, then the game is a pure conflict game such as the one represented in Fig. 3(b). Nevertheless, regardless of John's preferences, this situation is intuitively best characterized as a mixed-motive game: On the one hand, the game involves a common interest: John and*

Blonde want to meet with each other. On the other hand, the game also involves a conflict: John wants to meet with Brunette but Blonde wants to prevent this meeting.

The Example 3 illustrates the problem with the payoff-based classification of games: actual complex motives of players are aggregated into a single (artificially constructed) motive – payoff maximization. As a result, a game involving elements of both conflict and common interest may sometimes appear as a game of pure conflict and at other times as a game of pure common interest, depending on which motive "prevails". Hence, for more adequate classification of games it is necessary to disaggregate players' payoffs and uncover their various motives. Such disaggregation may also be useful for practical purposes as shown by the following modification of Example 3.

EXAMPLE 4 (*Dating games with mixed and opposed motives*): *First, consider the version of the dating game of Example 3 in which John wants to meet with both Blonde and Brunette but he prefers meeting with the latter. Call this version of the dating game Version 1. Next, consider a different version of the dating game, in which John wants to avoid Blonde while everything else remains the same. Call this version of the dating game Version 2. Both versions of the game can be represented as a pure conflict game depicted in Fig. 1(b). However, only in the Version 2 are the interests of John and Blonde directly opposed: Blonde wants to meet with John but John wants to avoid Blond; John wants to meet with Brunette but Brunette wants to prevent John from this meeting.*

More importantly, in the Version 1, Blonde can turn this game into a pure common-interest game by disposing of Brunette. This solution cannot be inferred from the standard representation of the game which does not provide information about players' goals.

Furthermore, Blonde of the second version is unable to turn the game into the one of pure common interest: if she disposes of Brunette, the game continues to be a pure conflict game. This difference between the Versions 1 and 2 again cannot be inferred from the standard representation.

My approach which explicitly models players' goals, allows for a more adequate definition of pure conflict, pure common-interest and mixed-motive games, which involves the standard definition as a special case.

DEFINITION 5: *Let* $\Gamma = \langle N, (A_i), (G_i), (S_i), \Omega, p, (c_i), \succsim_i \rangle$.

(a) Γ *is a* ***pure common-interest game****, if for every* $i, j \in N$ *and every* $(s, \omega) \in S \times \Omega$,

$c_i(s, \omega) = c_j(s, \omega)$.

(b) Let Γ *be such that* $|N| = 2$ *;* Γ *is a* ***pure conflict game****, if for every* $(s, \omega) \in S \times \Omega$*, either*

$c_1(s, \omega) = (0, ..., 0)$ *and* $c_2(s, \omega) = (1, ..., 1)$*, or* $c_1(s, \omega) = (1, ..., 1)$ *and* $c_2(s, \omega) = (0, ..., 0)$.

(c) Γ *is a* ***mixed-motive game****, if it is neither pure common-interest nor pure conflict game.*

To illustrate the Definition 5, I consider again the Version 1 of a dating game of Examples 3 and 4 and I model it as a strategic game with goal-oriented strategies.

EXAMPLE 5 (*A dating game with mixed motives as a strategic game with goal-oriented strategies*): *Let* $N = \{John, Blonde\}$, $A_J = A_B = \{X, Y\}$,

$G_J = \{Meet\ with\ Blonde,\ Meet\ with\ Brunette\}$,

$G_B = \{Meet\ with\ John,\ Prevent\ John\ from\ meeting\ Brunette\}$,

$S_J = \{(X, Blonde, Brunette), (Y, Blonde, Brunette)\}$*, and*

$S_B = \{(X, John, Prevent\ Brunette), (Y, John, Prevent\ Brunette)\}$. *Assume that there is only one state of nature. Probabilities of success are shown in the Fig. 4(a).*

	(*X*, *John*, *Prevent Brunette*)	(*X*, *John*, *Prevent Brunette*)
(*X*, *Blonde*, *Brunette*)	(1, 0), (1, 1)	(0, 1), (0, 0)
(*Y*, *Blonde*, *Brunette*)	(0, 1), (0, 0)	(1, 0), (1, 1)

Fig. 4: *A dating game as a strategic game with goal-oriented strategies*

According to the Definition 5, the dating game of the Example 5 is mixed-motive. Under what condition does the definition of pure common-interest, pure conflict, and mixed-motive game proposed here corresponds to the standard definition? First note, that if a game is of pure common interest (pure conflict) according to the Definition 2, then this game is also of pure common interest (pure conflict) according to the standard definition. Nevertheless, as shown in the Examples 3 and 5, the reverse is not generally true.

Above I have argued that the problem with the standard definition is that payoffs do not provide information about players' underlying goals. Hence, if each player has only one goal in mind and there is only one state of nature, then no information is lost if these goals are not explicitly specified. In this case, the classification of games as pure common-interest, pure conflict, and mixed-motive is always the same whether the standard or the new definition is used. Each player's preferences can be represented by their success functions; a game is then of pure common interest in the sense of the Definition 5 if and only if payoffs are perfectly positively correlated; a game is of pure conflict in the sense of Definition 5 if and only if payoffs are perfectly negatively correlated.[8]

3.2 Goal-orientedness and unification of behavioral sciences

The notion of goal-oriented strategies may contribute to the research program of unification of behavioral sciences. For some authors advancing this program, maximizing behavior has a place in the unified theory (Gintis 2009), for others it does not. For example, Vanberg (2002; 2004) argues that the principle of utility maximization should be replaced with Mayr's (1988; 1992) idea of goal-oriented program-based behavior (see also Conte and Castelfranchi 1995 for a similar argument).

This article shows that these two approaches to behavior are in fact compatible. The concept of goal-oriented strategy does not necessarily presuppose that individuals choose their goals consciously; nothing prevents one from interpreting goal-oriented strategies as programs. The preference relation defined on the bundles of probabilities over player's goals merely reflects unequal importance of various goals to the player (who may be a living or a non-living system) and is open to various interpretations. It may reflect player's subjective preferences (if it is a human being), contributions of player's goals to its fitness (if it is an organism), preferences of the engineer who designed the player (if it is a machine), or any other criterion. If a player has more than one goal, a model of behavior needs to incorporate some sort of

[8] The way how players' goals are defined, requires some attention. For instance, John of the dating game considered above, may want to meet with both Blonde and Brunette but perhaps he does not want to meet with both of them at the same time. Therefore, we may have $(1,0) \succ_J (1,1)$. If such an outcome is feasible, then John's goals can be more conveniently defined as "Meet with Blonde alone" and "Meet with Brunette alone". The general point is that the specification of goals has to be sufficiently detailed, so that all characteristics relevant to players' evaluations are included.

preference relation which would describe how agents resolve trade-offs between competing goals. Therefore, concept of goal-orientedness usually needs to be complemented with some sort of preference ranking. But the reverse is also sometimes true: preference ranking can be in some situations fruitfully complemented with the concept of goal-orientedness.

To see this, note that the standard game-theoretic assumptions allow only for one method of adjusting strategies to the environment at a time: either natural selection (if payoffs are interpreted as players' fitness) or learning and reasoning (if payoffs represent subjective preferences). The distinction used in this article between actions (means) and goals enables an analysis of various adjustment processes of adaptation within one framework. For instance, it can be assumed that natural selection operates on the set of goals (i.e. it determines the payoffs) and learning and reasoning operates on the level of adjustment of actions to given goals (i.e. it is concerned whether a particular strategy was successful in achieving a given goal or not) (El Mouden et al. 2012). The model of games with goal-oriented strategies can thus provide a useful link between social sciences and biology.

4 Concluding remarks

The model of games with goal-oriented strategies can be elaborated in several directions. Firstly, the idea of goal-oriented strategy can be straightforwardly extended to more complicated games than those considered in this article. Secondly, the concept of perfectly successful goal-oriented strategy seems suitable for analyzing situations in which players do not have common knowledge of the payoffs and the structure of the game: note that players can determine whether their strategies were successful or not without knowing the outcome of the play. Consequently, the model of games with goal-oriented strategies allow for analyzing various adjustment processes in strategic interactions using the probability of a player's goals success as a criterion of their performance. This line of development is of special interest, since a satisfactory theory of learning which would be applicable to both humans and non-humans is still lacking.

5 References

Ashby, Ross W. 1957. *An Introduction to Cybernetics*. London: Chapman & Hall.

Bacharach, Michael. 2006. *Beyond Individual Choice: Teams and Frames in Game Theory*. Edited by Natalie Gold and Robert Sugden. Princeton and Oxford: Princeton University Press.

Basu, Kaushik. 1994. "The Traveler's Dilemma: Paradoxes of Rationality in Game Theory." *The American Economic Review* 84 (2): 391–95.

Beard, T. Randolph, and Richard O. Beil Jr. 1994. "Do People Rely on the Self-Interested Maximization of Others? An Experimental Test." *Management Science* 40 (2): 252–62.

Becker, Gary S. 1998. *Accounting for Tastes*. Cambridge: Harvard University Press.

Bertalanffy, Ludwig von. 1968. *General System Theory*. New York: George Braziller.

Bicchieri, Cristina. 2005. *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Castelfranchi, Cristiano, and Rosaria Conte. 1998. "Limits of Economic and Strategic Rationality for Agents and MA Systems." *Robotics and Autonomous Systems* 24 (3–4): 127–39.

Colman, Andrew M. 1995. *Game Theory and Its Applications in the Social and Biological Sciences*. Oxford: Butterworth-Heinemann.

Conte, Rosaria, and Cristiano Castelfranchi. 1995. *Cognitive and Social Action*. London: University College London Press.

Dietrich, Franz, and Christian List. 2013a. "A Reason-Based Theory of Rational Choice." *Noûs* 47 (1): 104–34.

———. 2013b. "Where Do Preferences Come From?" *International Journal of Game Theory* 42 (3): 613–37.

El Mouden, Claire, Maxwell Burton-Chellew, Andy Gardner, and Stuart A. West. 2012. "What Do Humans Maximize?" In *Evolution and Rationality*, edited by Samir Okasha and Ken Binmore. Oxford: Cambridge University Press.

Gintis, Herbert. 2009. *The Bounds of Reason: Game Theory and the Unification of the Behavioral Sciences*. Princeton: Princeton University Press.

Goeree, Jacob K., and Charles A. Holt. 2001. "Ten Little Treasures of Game Theory and Ten Intuitive Contradictions." *The American Economic Review* 91 (5): 1402–22.

Hayek, F. A. 1937. "Economics and Knowledge." *Economica*, New Series, 4 (13): 33–54.

Hudík, Marek. 2014. "A Preference Change or a Perception Change? A Comment on Dietrich and List." *International Journal of Game Theory* 44 (2): 425–31.

———. 2015. "Goals, Payoffs, and Classification of Games." http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2632152.

Kohlberg, Elon, and Jean-Francois Mertens. 1986. "On the Strategic Stability of Equilibria." *Econometrica* 54 (5): 1003–37.

List, John A. 2006. "'Friend or Foe?' A Natural Experiment of the Prisoner's Dilemma." *The Review of Economics and Statistics* 88 (3): 463–71.

Locke, Edwin A., and Gary P. Latham. 2002. "Building a Practically Useful Theory of Goal Setting and Task Motivation: A 35-Year Odyssey." *American Psychologist* 57 (9): 705–17.

———. , eds. 2013. *New Developments in Goal Setting and Task Performance*. New York and London: Routledge.

Mayr, Ernst. 1988. *Toward a New Philosophy of Biology: Observations of an Evolutionist*. Harvard University Press.

———. 1992. “The Idea of Teleology.” *Journal of the History of Ideas* 53 (1): 117–35.

McKelvey, Richard D., and Thomas R. Palfrey. 1992. “An Experimental Study of the Centipede Game.” *Econometrica* 60 (4): 803–36.

Palacios-Huerta, Ignacio, and Oscar Volij. 2009. “Field Centipedes.” *The American Economic Review* 99 (4): 1619–35.

Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow. 1943. “Behavior, Purpose and Teleology.” *Philosophy of Science* 10 (1): 18–24.

Rosenthal, Robert W. 1981. “Games of Perfect Information, Predatory Pricing and the Chain-Store Paradox.” *Journal of Economic Theory* 25 (1): 92–100.

Schelling, Thomas C. 1980. *The Strategy of Conflict*. Cambridge: Harvard University Press.

Sugden, Robert. 2000. “Team Preferences.” *Economics and Philosophy* 16 (02): 175–204.

Vanberg, Viktor J. 2002. “Rational Choice vs. Program-Based Behavior Alternative Theoretical Approaches and Their Relevance for the Study of Institutions.” *Rationality and Society* 14 (1): 7–54.

———. 2004. “The Rationality Postulate in Economics: Its Ambiguity, Its Deficiency and Its Evolutionary Alternative.” *Journal of Economic Methodology* 11 (1): 1–29.