Article

The Role of Implicit Motives in Strategic Decision-Making: Computational Models of Motivated Learning and the Evolution of Motivated Agents

Kathryn Merrick

School of Engineering and Information Technology, University of New South Wales, Canberra 2600, Australia; E-Mail: k.merrick@adfa.edu.au; Tel.: +61-2-6268-8023; Fax: +61-2-6268-8443

Abstract: Individual behavioral differences in humans have been linked to measurable differences in their mental activities, including differences in their implicit motives. In humans, individual differences in the strength of motives such as power, achievement and affiliation have been shown to have a significant impact on behavior in social dilemma games and during other kinds of strategic interactions. This paper presents agent-based computational models of power-, achievement- and affiliation-motivated individuals engaged in game-play. The first model captures learning by motivated agents during strategic interactions. The second model captures the evolution of a society of motivated agents. It is demonstrated that misperception, when it is a result of motivation, causes agents with different motives to play a given game differently. When motivated agents who misperceive a game are present in a population, higher explicit payoff can result for the population as a whole. The implications of these results are discussed, both for modeling human behavior and for designing artificial agents with certain salient behavioral characteristics.

Keywords: motivation; game theory; learning; evolution

1. Introduction

Enduring motive dispositions, or "implicit motives" are preferences for certain kinds of incentives that are acquired in early childhood [1]. Incentives are situational characteristics associated with possible satisfaction of a motive. Incentives can be either implicit or explicit. Examples of implicit

incentives include challenges to personal control in a performance situation (incentive for achievement), opportunities for social closeness (incentive for affiliation) or opportunities for social control (incentive for power). In humans, differences in implicit motives have also been linked to differences in preferences for explicit incentives such as money, points or "payoff" in a game [2–5] or during other kinds of strategic interactions [6]. Computational motivation has emerged as an area of interest among artificial intelligence [7,8] and robotics [9–11] researchers, as a mechanism for enabling autonomous mental development [12] or modelling human-like decision-making [8,13]. This paper falls in this latter category.

The contribution of this paper is two agent-based approaches to modelling the behavior of power-, achievement- and affiliation-motivated individuals engaged in game-play: a motivated learning agent model for controlling the behavior of individual agents, and an evolutionary model for controlling the proportions of agents with different types of motives in a society of motivated agents. The motivated learning model is studied in the context of two well-known, 2 × 2 mixed-motive games: the prisoner's dilemma (PD) game, and the snowdrift game. Analysis shows that subjectively rational agents that act to satisfy their implicit motives perceive games that have different Nash Equilibrium (NE) points to the original game. The implications of this result are demonstrated in interactions between computational agents with different motives. We compare the behavioral characteristics of the agents with those observed in humans with the corresponding motives, and discuss some of the qualitative similarities achieved by the computational model.

The motivated evolutionary model is used to examine the question of whether there is an evolutionary benefit of motivation-based misperception in computational settings. Motivated agents are studied in two multiplayer games: a common pool resource (CPR) game, and the canonical hawk-dove game. It is demonstrated that motivated agents who misperceive a game can form stable sub-populations and higher explicit payoff can result for the population as a whole. The implications of these results are discussed, both for modelling human behavior and for designing artificial agents with certain salient behavioral characteristics.

The remainder of this section is organized as follows. First, we briefly review the literature of incentive-based motivation theories in Section 1.1. A summary is given of the different behavioral characteristics that have been observed in humans with different dominant motives, and thus different implicit motive profiles. This is used as the basis for qualitative evaluation of the models in Section 3. The assumptions of this study and related work are outlined in Section 1.2. Section 2 presents the new agent models, which are then examined theoretically and experimentally in Section 3.

1.1. Achievement, Affiliation and Power Motivation

Theories of achievement, affiliation and power motivation have been considered particularly influential by psychologists [1]. They form the basis of theories such as the three needs theory [14] and three factor theory [15]. This paper is specifically concerned with modelling the influence of different dominant motives (specifically either achievement, affiliation or power motivation) during strategic decision-making. This section briefly reviews some of the salient characteristics associated with each motive, which are also summarized in Table 1.

Table 1. Characteristics that may be observed in individuals with a given dominant motive [1,14].

Dominant Motive	Possible Behavioral Characteristics
Achievement	• Prefers moderately challenging goals • Willing to take calculated risks • Likes regular feedback • Often likes to work alone
Affiliation	• Wants to belong to a group • Wants to be liked • Prefers collaboration over competition • Does not like high risk or uncertainty
Power	• Wants to control and influence others • Likes to win • Likes competition • Likes status and recognition

Achievement motivation drives humans to strive for excellence by improving on personal and societal standards of performance. A number of models exist, including Atkinson's Risk-Taking Model (RTM) [16] and more recent work that has examined achievement motivation from an approach-avoidance perspective [17]. The aspect of achievement motivation of interest in this paper is the hypothesis that success-motivated individuals perceive an inverse linear relationship between incentive and probability of success [6,18]. They tend to favor goals or actions with moderate incentives, which can be interpreted as indicating a moderate probability of success, calculated risk, or moderate difficulty. They are often content to work alone to achieved these goals.

Approach-avoidance motivation has also been studied into the social domain [19]. In this domain it is used to model the differences in goals concerned with positive social outcomes—such as affiliation and intimacy—and goals concerned with negative social outcomes—such as rejection and conflict [20]. It is understood that the idea of approach-avoidance motivation, along with the concepts of incentive and probability of success, are particularly important not only in achievement motivation, but also affiliation, power and other forms of motivation [1].

Affiliation refers to a class of social interactions that seek contact with formerly unknown or little known individuals and maintain contact with those individuals in a manner that both parties experience as satisfying, stimulating and enriching [1]. The need for affiliation is activated when an individual comes into contact with another unknown or little known individual. While theories of affiliation have not been developed mathematically to the extent of the RTM, affiliation can be considered from the perspective of incentive and probability of success [1]. In contrast to achievement-motivated individuals, individuals high in affiliation motivation may select goals with a higher probability of success and/or lower incentive. This, often counter-intuitive preference, can be understood as avoiding risk, public competition or conflict. Rather, affiliation motived individuals prefer to "belong to a group". Affiliation motivation is considered an important balance to power motivation [1].

Power can be described as a domain-specific relationship between two individuals, characterized by the asymmetric distribution of social competence, access to resources or social status [1]. Power is manifested by unilateral behavioral control and can occur in a number of different ways. Types of power include reward power, coercive power, legitimate power, referent power, expert power and

informational power. As with affiliation, power motivation can be considered with respect to incentive and probability of success. Specifically, there is evidence to indicate that the strength of satisfaction of the power motive depends solely on incentive and is unaffected by the probability of success [21] or risk. Power motivated individuals select high-incentive goals, as achieving these goals gives them significant control of the resources and reinforcers of others. Power motivated individuals like competition and like to win.

McClelland [14] writes that, regardless of gender, culture or age, an individual's implicit motive profile tends to have a dominating motivational driver. That is, one of the three motives discussed above will have a stronger influence on decision-making than the other two, but the individual will not be conscious of this. The dominant motive might be a result of cultural or life experiences and results in distinct individual characteristics, some of which are summarized in Table 1.

Hybrid profiles of power, affiliation and achievement motivation have also been associated with distinct individual characteristics. For example, there appears to be a relationship between certain combinations of dominant and non-dominant motives and the emergence of leadership abilities in an individual [22]. However, this paper focuses on modeling agents with a single dominant motive.

1.2. Assumptions and Related Work

Previous work has modeled incentive-based profiles of power, achievement and affiliation motivation computationally for agents making one-off decisions [13]. For example, Figure 1 shows a possible computational motive profile as a function of incentive. Motivation is modeled as the sum of three sigmoid curves for achievement, affiliation and power motivation. The height of each sigmoid curve corresponds to the strength of the motive. The strongest motive is the dominant motive. The dominant motive in the profile in Figure 1 is thus power motivation.
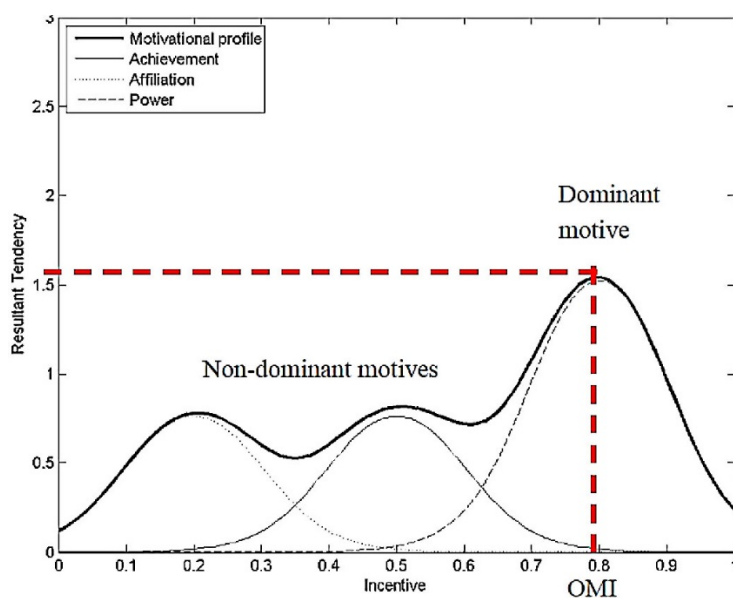

Figure 1. A computational motive-profile. The resultant tendency for action is highest for incentive of 0.8, which is the optimally motivating incentive (OMI) for this agent. This agent may be qualitatively classified as "power-motivated" as its OMI is relatively high on the zero-to-one scale for scale for incentive. Power motivation is the dominant motive.

Later work simplifies this kind of model for game-theoretic settings using the concept of an optimally motivating incentive (OMI) [8,23]. The OMI is the incentive value that maximizes a motivation curve. Assuming a fixed range for incentive, agents are qualitatively classified as power-, achievement- or affiliation-motivated if the incentive value that optimizes (maximizes) their motivation curve is "high", "moderate" or "low" respectively in the range. This paper builds on the OMI approach, proposing agent models that incorporate OMIs to bias agents' perception of a game. These models are presented in Section 2. We first outline the assumptions of these models, and how they differ to previous work.

Two key assumptions in this paper are (1) that game theoretic "payoff" is used to represent incentive and (2) that individual agents are subjectively rational in their quest for this payoff. The first assumption means that this paper focuses on the influence of implicit motives when judging explicit incentives.

The second assumption means that different agents may perceive the same explicit incentive (payoff) differently as a result of having different OMIs. They execute behaviors B_t so as to maximise their own perceived, subjective incentive. We define the subjective incentive of an agent as follows:

$$\hat{I}_t = V^{max} - |V_t - \Omega| \tag{1}$$

where V^{max} is the maximum possible explicit incentive (payoff) and V_t is the explicit incentive received for executing behavior B_{t-1}. Ω is the agent's OMI, which is fixed for the lifetime of the agent. The logic behind this definition is that an individual's subjective incentive is higher if there is a smaller difference between the explicit incentive and their OMI. That is, if $|V_t - \Omega|$ is small. Subtraction of this term from V^{max} serves to normalize the resulting value to the range $(0, V^{max})$.

Other studies have considered the case where individuals are subjectively irrational [24,25]. In this paper, the assumption of subjective rationality hinges on the link between motivation and action. In particular, extensive experimental evidence indicates that individuals will act, apparently sub-optimally according to some objective function, to fulfil their (subjective) motives [1].

This paper further assumes (3) there is no communication between agents; and (4) the motivations of other agents are not observable. Non-communication between agents is a standard game theoretic assumption and we adopt it in this paper. Specifically we mean that agents do not communicate any indication of the decision they will make *before* they make the decision. The decision may of course be communicated once made.

The non-observability of motivations is assumed based on the difficulty of identifying an individual's implicit motive profile. This assumption differentiates our work from other work on altered perception or misperception where the perceptions of others are assumed to be observable [24]. We do not model perception of a game in terms of expectations about other players' strategies. Such approaches have been taken in other related game [25], metagame [26] and hypergame [27] theory research concerning misperception or the evolution of preferences. Rather we model perception as a process of transforming one game into another.

Other related work includes game theoretic frameworks for personality traits [28], reciprocity [29,30], aspiration models [31] and work on misperception in a game theoretic setting [32]. The work in this paper differs from these works through its specific focus on motivation rather than other kinds of personality traits, reciprocity, aspirations or misperception as a result of other factors.

Various techniques have also been explored previously for modelling agent learning during game theoretic interactions. These include learning from fictitious play [33], memory bounded learning [34] to compensate for non-stationary strategies by the opposition and cultural learning through reinforcement and replicator dynamics [35]. The models in this paper augment this latter approach to develop motivated learning agents.

Yet, other work has considered the role of evolution in motivated agents, including the evolution of internal reinforcers [36,37]. Our work differs from this as it focuses on the evolution of a society of agents with different motivational preferences, rather than the evolution of reward signals.

2. Materials and Method

This section presents the new agent algorithms for motivated learning (Section 2.1) and evolving the proportions of different motivated agents in a population (Section 2.2). The first algorithm combines the concept of an OMI and motivated perception using Equation (1) with Cross learning formalized for two player games [35]. The second combines an OMI with a replicator equation for evolution.

2.1. Motivated Learning Agents

The motivated learning agent algorithm is shown in Algorithm 1. The algorithm is designed for a motivated agent interacting in a 2×2 game, that is, a two-player game in which each player has the choice of two behaviors B^C and B^D. B^C is understood to be a "cooperative" behavior, while B^D is defined as a refusal to cooperate (also called defecting). The precise definition of cooperate or defect depends on the precise nature of the game. Some specific examples are given in Section 3, but in general, this paper focuses on mixed-motive games **W** of the form:

$$\mathbf{W} = \begin{bmatrix} P & T \\ S & R \end{bmatrix} \tag{2}$$

R denotes the "reward" for mutual cooperation. P denotes the "punishment" for mutual defection. T represents the "temptation" to defect, while S is the sucker's payoff for choosing to cooperate when the other player chooses to defect.

Algorithm 1. Algorithm for a motivated learning agent.

---
1. Initialize game world **W** of form in Equation (2), and identify V^{max}
2. Initialize agent's optimally motivating incentive Ω and $P(B_0 = B^C)$ and $P(B_0 = B^D)$
3. Repeat:
4. If $t > 0$
5. Receive payoff V_t for previously executed behavior B_{t-1}
6. Compute subjective incentive $\hat{I}_t$ using Equation (1)
7. Compute $P(B_t = B^C)$ and $P(B_t = B^D)$ using Equation (3)
8. Generate a random number *r* from (0, 1)
9. If $r < P(B_t = B^C)$ select behavior $B_t = B^C$ else select behavior $B_t = B^D$
10. Store and execute B_t

This game is initialized in line 1 of Algorithm 1. The agent is then initialized in line 2 with an OMI Ω and initial probabilities $P(B_0 = B^C)$ and $P(B_0 = B^D)$. $P(B_0 = B^C)$ is the probability that the

behavior B^C is executed at time $t = 0$. $P(B_0 = B^C)$ is chosen at random from a uniform distribution and $P(B_0 = B^D) = 1 - P(B_0 = B^C)$.

When $t = 0$, the agent selects an action probabilistically based on $P(B_0 = B^C)$ as shown in lines 8–9. In each subsequent iteration, the agent receives a payoff V_t for its previously executed behavior (line 5) then computes a subjective incentive value based on its individual OMI (line 6).

Next, the agent updates its probabilities of choosing B^C and B^D (line 7). Cross learning formalized for two player games [35] is used to model this cultural learning as follows:

$$P(B_t = B^i) = \begin{cases} [1 - \alpha \hat{I}_t] P(B_{t-1} = B^i) + \alpha \hat{I}_t \; if \; B_{t-1} = B^i \\ [1 - \alpha \hat{I}_t] P(B_{t-1} = B^i) \; otherwise \end{cases} \tag{3}$$

where *i* is either *C* or *D* and α is the learning rate.

The agent chooses its next behavior probabilistically (lines 8–9) based on the updated values of $P(B_0 = B^C)$ and $P(B_0 = B^D)$. The chosen behavior is then stored and executed (line 10). The intelligence loop in lines 4–10 is repeated as long as the agent is alive and/or has an opponent against which to play.

As we will see in the analysis in Section 3, motivated learning agents will adapt their strategy over time, influenced by their own OMI and their opponent's behavior. However, as they age, they will eventually converge on a stable strategy. If we would like to model the introduction of new generations of agents to a society of game-playing agents, then another algorithm is required. This is the topic of the next section.

2.2. Evolution of Motivated Agents

We model a society of *n* game-playing agents using the standard assumption that every agent in each new generation of agents engages in a two-player game with every other agent in its generation. The *n*-player game is thus a compound two player game. In such a game, the total payoff to a single agent when all other agents choose the same behavior as each other is either $T(n - 1)$, $S(n - 1)$, $R(n - 1)$ or $P(n - 1)$.

To construct the algorithm for the evolution of a society of motivated agents (Algorithm 2) we first select payoff constants *T*, *R*, *S* and *P* (line 1) and use these to construct a game world **W** comprised of *j* different types of motivated agents. An agent type is defined by its OMI. The total number of different types of agents (*i.e.*, the maximum number of different OMIs among agents in the society) is *J*. The game world **W** in which motivated agents evolve, is represented by a 2*J* by 2*J* matrix concatenating the perceived games of each of the *J* types of agents vertically, and then repeating this horizontally *J* times. These games are constructed using Equations (5)–(8), which are derived from Equation (1) using the assumption that $V^{max} = T(n - 1)$.

$$\mathbf{W} = \begin{bmatrix} \hat{P}^1 & \hat{T}^1 & \hat{P}^1 & \hat{T}^1 \\ \hat{S}^1 & \hat{R}^1 & \cdots & \hat{S}^1 & \hat{R}^1 \\ & & \cdot & & \\ \hat{P}^j & \hat{T}^j & \hat{P}^j & \hat{T}^j \\ \hat{S}^j & \hat{R}^j & \cdots & \hat{S}^j & \hat{R}^j \\ & & \cdot & & \\ \hat{P}^J & \hat{T}^J & \hat{P}^J & \hat{T}^J \\ \hat{S}^J & \hat{R}^J & \cdots & \hat{S}^J & \hat{R}^J \end{bmatrix} \qquad (4)$$

$$\hat{T}^j = T(n-1) - \mid T(n-1) - \Omega^j \mid \qquad (5)$$

$$\hat{R}^j = T(n-1) - \mid R(n-1) - \Omega^j \mid \qquad (6)$$

$$\hat{P}^j = T(n-1) - \mid P(n-1) - \Omega^j \mid \qquad (7)$$

$$\hat{S}^j = T(n-1) - \mid S(n-1) - \Omega^j \mid \qquad (8)$$

Algorithm 2. Algorithm for evolving the proportions of agents with different motives in a society of motivated agents.

---

1. Initialize T, R, P, S and matrix $\mathbf{W}$ of form in Equation (4)
2. Initialize $\mathbf{x}_0$ of form in Equation (9) and n.
3. Initialize a society $\mathbf{A}$ of n agents with correct proportions of each type according to $\mathbf{x}_0$.
4. Repeat:
5. Compute $\mathbf{x}_{t+1}$ using update in Equations (11) and (12)
6. Normalize $\mathbf{x}_{t+1}$ (see text for details)
7. Create a new generation of agents in society $\mathbf{A}$ (and remove the old generation) to reflect new proportions in $\mathbf{x}_{t+1}$

We then construct a vector $\mathbf{x}_t$ stipulating the proportions of each type of agent in the society, further broken down by the proportion at any time choosing B^C and B^D (line 2).

$$\mathbf{x}_t = \begin{bmatrix} F^1(B_t = B^C) \\ F^1(B_t = B^D) \\ \cdot \\ \cdot \\ F^j(B_t = B^C) \\ F^j(B_t = B^D) \\ \cdot \\ \cdot \\ F^J(B_t = B^C) \\ F^J(B_t = B^D) \end{bmatrix} \qquad (9)$$

$F^j(B_t = B^i)$ is the fraction of all agents who are of type j (with OMI Ω^j) and will choose B^i. This means that $\sum_i F^j(B_t = B^i)$ is the fraction of agents of type j. The probability of an agent of a given type choosing B^C is:

$$P^j(B_t = B^C) = \frac{F^j(B_t = B^C)}{\sum_i F^j(B_t = B^i)} \qquad (10)$$

and likewise for B^D.

A specific number of agents *n* is then chosen (line3) and a society **A** of *n* agents created such that the distribution of agent OMIs conforms to the proportions initialized in $\mathbf{x}_0$. These agents may be motivated learning agents, or indeed any other kind of motivated agent in which motivation and behavior are governed by an OMI.

The evolutionary process is governed by the replicator equation:

$$\dot{\mathbf{x}}_{t+1} = \mathbf{Q}\mathbf{x}_t[\mathbf{W}\mathbf{x}_t - \mathbf{x}_t{}^{\mathrm{T}}\mathbf{W}\mathbf{x}_t] \tag{11}$$

Q is a matrix of transition probabilities for the mutation of agent type *j* to agent type *k*. When **Q** is the identity matrix, Equation (11) models evolution without mutation. The Euler method is applied to simulate the process of evolution (line 5):

$$\mathbf{x}_{t+1} = \mathbf{x}_t + h\dot{\mathbf{x}}_t \tag{12}$$

where *h* is a step size parameter that controls the rate of change of the composition of the population. Smaller step sizes result in slower changes to the composition of the population, and a generally more stable population. Large step sizes can produce large fluctuations in the proportions of the different types of agents. $\mathbf{x}_{t+1}$ must be normalised by zeroing any negative fractions and ensuring that $\sum_j \sum_i F^j(B_t = B^i) = 1$ (line 6). Finally the composition of **A** is updated by creating a new generation of agents with the appropriate new distribution of OMIs (line 7).

Evolutionary dynamics predict the equilibrium outcomes of a multi-agent system when the survival of agents in iterative game-play is determined by their fitness. Using this approach, the proportion of a given type of agent playing B^C or B^D will increase if the subjective value of incentives earned by that type of agent is greater than the average subjective value perceived by the population. Conversely the proportion of a given type of agent will decrease if the subjective value of incentive earned is less than the average perceived by the population. Agent types with that perceive precisely the average subjective incentive will neither increase nor decrease in number, unless a mutation occurs. Agents are thus considered to be fitter if they satisfy their own motives, regardless of whether they receive the highest explicit incentive as defined in the original game.

3. Results and Discussion

This section examines learning by, and evolution of, motivated agents in two settings. The first setting, examined in Section 3.1, is a two-player PD game and its *n*-player compound, a CPR game. The second setting, examined in Section 3.2, is a two-player snowdrift game and its *n*-player compound, the hawk-dove game. In each setting, a theoretical examination of learning and evolutionary behavior is presented first, followed by empirical examinations of learning in the two-player game and evolution in the corresponding *n*-player compound game. The theoretical analysis demonstrates how different motivated agents perceive a game in terms of the subjective incentives they perceive. The empirical investigations reveal what happens when agents with different motives actually play the game. A brief demonstration in Appendix B completes the progression from theory to application.

3.1. The Prisoners' Dilemma and Common Pool Resource Games

The two-player PD game [38] derives its name from a hypothetical strategic interaction in which two prisoners are held in separate cells and cannot communicate with each other. The police have

insufficient evidence for a conviction unless at least one prisoner discloses certain information. Each prisoner has a choice between concealing the information (B^C) or disclosing it (B^D). If both conceal, both with be acquitted and receive a payoff of *R*. This is denoted the (B^C, B^C) outcome. If both disclose, both will be convicted and receive minor punishments *P*. This is denoted the (B^D, B^D) outcome. If only one prisoner discloses information he will be acquitted and also receive a reward *T* for his information. The prisoner who conceals information unilaterally will receive a heavy punishment *S*. Thus, in a PD game it is assumed that $T > R > P > S$.

The compound *n*PD version of this game has been used as a model for arms races, voluntary wage restraint and the iconic "tragedy of the commons" as well as for conservation of scarce CPRs [26]. The dilemma confronting an individual citizen faced with a scarce CPR—such as energy or water for example—is whether to exercise restraint in resource use (B^C) or exploit the restraint of others (B^D). An individual benefits from restraint only if a large proportion of others also exercise restraint. However, if this condition holds, it appears to become individually rational to ignore the call for restraint. In practice this does not happen all the time and the following theoretical results offer insight into why this is the case in a society of motivated agents. The theoretical analysis here is done for the general case of *n* players. When *n* = 2 we have the special case of a two-player PD game. We use the two-player game in an empirical study of motivated learning in Section 3.1.2 and the *n* player game to study evolution of motivated agents in Section 3.1.3.

3.1.1. Theoretical Results

We consider three cases for theoretical analysis corresponding to perception caused by power, achievement and affiliation motivation. We use the following mapping of OMI values to motivation types [23]:

- Power-motivated agents have $T(n-1) > \Omega^j > \frac{1}{2}(T+P)(n-1)$
- Achievement-motivated agents have $\frac{1}{2}(T+P)(n-1) > \Omega^j > \frac{1}{2}(R+S)(n-1)$
- Affiliation-motivated agents have $\frac{1}{2}(R+P)(n-1) > \Omega^j > S(n-1)$

This mapping implies that power-motivated individuals prefer "high" incentives; achievement-motivated individuals prefer "moderate" incentives and affiliation-motivated individuals prefer "low" incentives. Furthermore, this mapping of motivation types to incentive values implies that "high", "moderate" and "low" can change depending on the distribution of values *T*, *R*, *S* and *P*. The division of OMIs into "high", "medium" and "low" ranges is informed by the literature on incentive-based motives discussed in Section 1.1, but represents a stronger mathematical assumption than has previously been made by psychologists. Specifically, while previous experiments with humans have provided evidence that individuals with different dominant motives prefer different goals, these experiments have not attempted to define precise value ranges for the incentives that will be preferred by different individuals.

Power-Motivated Perception

For power-motivated agents we assume $T(n-1) > \Omega^j > \frac{1}{2}(T+P)(n-1)$. Two different games may be perceived by power-motivated agents, depending on whether Ω^j is closer to $T(n-1)$ or $R(n-1)$.

First, $T(n-1) > \Omega^j > \frac{1}{2}(T+R)(n-1)$ (*i.e.*, Ω^j closer to $T(n-1)$)) gives us the following transformation of the PD game using the Equations (5)–(8) and simplifying the absolute values:

$$\hat{T}^j = T(n-1) - (T(n-1) - \Omega^j) = \Omega^j \tag{13}$$

$$\hat{R}^j = T(n-1) - (\Omega^j - R(n-1)) = (T+R)(n-1) - \Omega^j \tag{14}$$

$$\hat{P}^j = T(n-1) - (\Omega^j - P(n-1)) = (T+P)(n-1) - \Omega^j \tag{15}$$

$$\hat{S}^j = T(n-1) - (\Omega^j - S(n-1)) = (T+S)(n-1) - \Omega^j \tag{16}$$

Theorem 1: When a player with $T(n-1) > \Omega^j > \frac{1}{2}(T+R)(n-1)$ perceives a nPD/CPR game **W** with $T > R > P > S$, the game they perceive is still a valid nPD/CPR game with $\hat{T}^j > \hat{R}^j > \hat{P}^j > \hat{S}^j$.

Proofs of the theorems in this section are given in Appendix A.

When $\frac{1}{2}(T+R)(n-1) > \Omega^j > \frac{1}{2}(T+P)(n-1)$ (*i.e.*, Ω^j closer to $R(n-1)$)), there is an additional transformation for the case when $\Omega^j < R(n-1)$:

$$\hat{R}^j = T(n-1) - (R(n-1) - \Omega^j) = (T-R)(n-1) + \Omega^j \tag{17}$$

Theorem 2: When a player with $\frac{1}{2}(T+R)(n-1) > \Omega^j > \frac{1}{2}(T+P)(n-1)$ perceives a nPD/CPR game **W** with $T > R > P > S$, the game they perceive will have $\hat{R}^j > \hat{T}^j > \hat{P}^j > \hat{S}^j$

Figure 2 visualizes the expected subjective incentive of games perceived by one power-motivated agent of type j given that a certain number of other agents $n^C < n$ play B^C. It is clear from the visualization that the strategy of always playing B^D dominates for $\frac{1}{2}(T+R)(n-1) < \Omega^j$ (see Figure 2a), but no longer dominates when $\frac{1}{2}(T+R)(n-1) > \Omega^j$ (see Figure 2b). The implication is that there are conditions under which a power-motivated agent will exhibit a preference for choosing B^C. Specifically, they may do so if they have a lower OMI in the power-motivated range, and if there is a sufficiently large proportion of other agents choosing B^C.
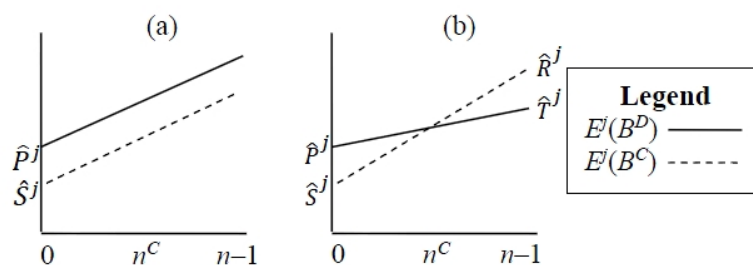

Figure 2. Visualization of the subjective incentive perceived by a power-motivated agent playing an explicit PD game according to (**a**) Theorem 1 and (**b**) Theorem 2.

Achievement-Motivated Perception

For achievement-motivated individuals we assume $\frac{1}{2}(T+P)(n-1) > \Omega^j > \frac{1}{2}(R+S)(n-1)$. We again consider two cases, depending on whether Ω^j falls closer to $R(n-1)$ or $P(n-1)$. When it falls closer to $R(n-1)$ we have:

Theorem 3: When a player with $\frac{1}{2}(T + P)(n - 1) > \Omega^j > \frac{1}{2}(R + P)(n - 1)$ perceives a nPD/CPR game **W** with $T > R > P > S$, the game they perceive is either:

$\hat{R}^j > \hat{P}^j > \hat{T}^j > \hat{S}^j$ if $\Omega^j > \frac{1}{2}(T + S)(n - 1)$ or

$\hat{R}^j > \hat{P}^j > \hat{S}^j > \hat{T}^j$ if $\Omega^j < \frac{1}{2}(T + S)(n - 1)$

Theorem 3 shows that the transformation of a game may depend on the distribution of incentive values as well as the value of Ω^j. $\hat{R}^j > \hat{P}^j > \hat{S}^j > \hat{T}^j$ results if the values of T, R, P and S have a highly non-linear distribution, with a much larger gap between T and R than between the other incentives. $\hat{R}^j > \hat{P}^j > \hat{T}^j > \hat{S}^j$ results from all other distributions, including linear distributions. Figure 3a,b visualize the structure of the perceived games for each of the cases in Theorem 3. In both games there are two pure strategy equilibria: "always play B^C" or "always play B^D". The former is more likely for players who perceive $\hat{R}^j > \hat{P}^j > \hat{S}^j > \hat{T}^j$. These games also both have a mixed strategy equilibrium, where either some players play B^C and some play B^D, or players alternate between the two pure strategies with some probability.
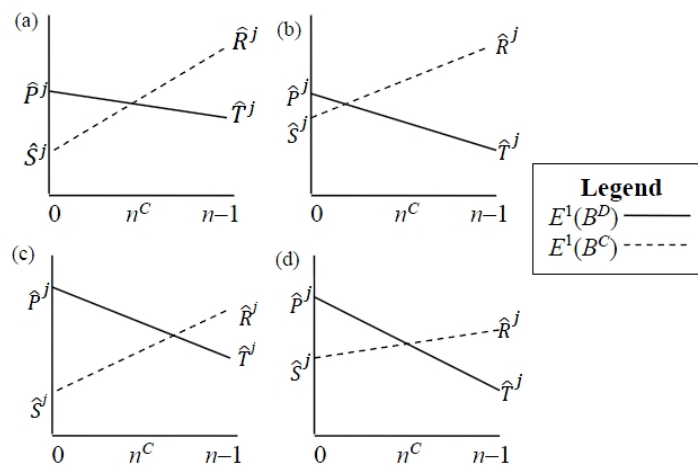

Figure 3. Visualization of the subjective incentive perceived by achievement-motivated agents playing an explicit PD game according to (**a,b**) Theorem 3 and (**c,d**) Theorem 4.

If Ω^j falls closer to $P(n - 1)$, the transformations in Equations (13) and (15)–(17) apply, with the addition of Equation (18) for the case when $\Omega^j < P(n - 1)$:

$$\hat{P}^j = T(n - 1) - (P(n - 1) - \Omega^j) = (T - P)(n - 1) + \Omega^j \tag{18}$$

Theorem 4: When a player with $\frac{1}{2}(R + P)(n - 1) > \Omega^j > \frac{1}{2}(R + S)(n - 1)$ plays a nPD/CPR game **W** with $T > R > P > S$, the game they perceive is either:

$\hat{P}^j > \hat{R}^j > \hat{T}^j > \hat{S}^j$ if $\frac{1}{2}(T + S)(n - 1) < \Omega^j$ or

$\hat{P}^j > \hat{R}^j > \hat{S}^j > \hat{T}^j$ if $\frac{1}{2}(T + S)(n - 1) > \Omega^j$

Once again, $\hat{P}^j > \hat{R}^j > \hat{T}^j > \hat{S}^j$ can only occur for a highly non-linear payoff distribution. Figure 3c,d visualize the structure of the perceived games for each of the cases in Theorem 4. In both games there are still two pure strategy equilibria, and the possibility of a mixed strategy equilibrium.

Affiliation-Motivated Perception

For affiliation-motivated agents, we assume $\frac{1}{2}(R + S)(n - 1) > \Omega^j > S(n - 1)$. We again consider two cases, depending on whether Ω^j is closer to $P(n - 1)$ or $S(n - 1)$. First we consider the case where Ω^j is closer to $P(n - 1)$.

Theorem 5: When a player with $\frac{1}{2}(R + S)(n - 1) > \Omega^j > \frac{1}{2}(P + S)(n - 1)$ plays a nPD/CPR game **W** with $T > R > P > S$, the game they perceive has $\hat{P}^j > \hat{S}^j > \hat{R}^j > \hat{T}^j$.

If Ω^j is closer to $S(n - 1)$, then the transformations in Equations (13) and (16)–(18) apply.

Theorem 6: When a player with $\frac{1}{2}(P + S)(n - 1) > \Omega^j > S(n - 1)$ plays a nPD/CPR game **W** with $T > R > P > S$, the game they perceive has $\hat{S}^j > \hat{P}^j > \hat{R}^j > \hat{T}^j$.

Figure 4 visualizes the structure of the perceived games for the cases in Theorems 5 and 6. We see in Figure 4b that affiliation-motivated agents with a very low OMI will perceive a game in which the "always play B^C" strategy dominates the "always play B^D" strategy. This means that affiliation-motivated agents with very low OMI will always converge on the "always play B^C" strategy over time.
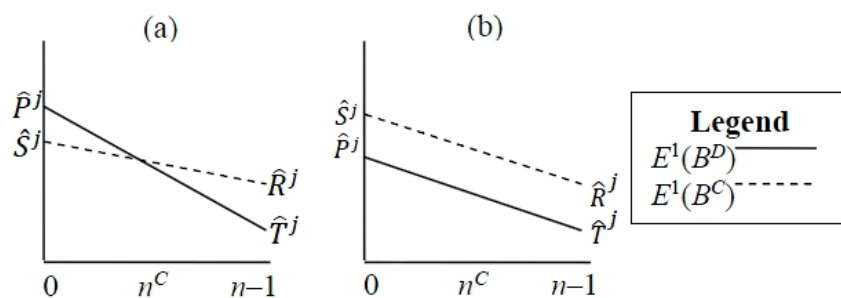

Figure 4. Visualization of the subjective incentive perceived by affiliation-motivated agents playing an explicit PD game according to (**a**) Theorem 5 and (**b**) Theorem 6.

The analysis above demonstrates how motivation results in a number of different perceived games with different subjective incentives. Some power-motivated individuals will still perceive a valid nPD/CPR game, but agents with other motives will perceive one of seven different games with different possible equilibria.

The next sections examine empirically, how these different types of agents learn during interactions with agents of other types and how evolution affects the proportions of different types of agents in a population.

3.1.2. Empirical Study of Motivated Learning in the Prisoners' Dilemma Game

In the experiments in this section, six sets of thirty pairs of agents A^1 and A^2 use Algorithm 1 to learn during 3000 iterations of a game. Each agent A^j has probabilities $P^j(B_t = B^i)$ and OMI Ω^j. $\alpha = 0.001$. The results in this section are presented visually in Figure 5 in terms of the change in $P^j(B_t = B^C)$ (and by implication, $P^j(B_t = B^D)$) as agents learn. The horizontal axis on each chart measures $P^1(B_t = B^C)$ and the vertical axis measures $P^2(B_t = B^C)$. Thus, a point in the upper right corner of each graph is indicative that both agents are choosing B^C. The other corners also have a prescribed meaning. Corners in which equilibria occur are labelled. Points closer to the center of the chart are indicative of

a more or less equal preference for B^C and B^D. The dominant motive of each agent is indicated in the sub-figure titles. These motives are:

- Power-motivated: $\Omega^j = 3.9$
- Achievement-motivated: $\Omega^j = 2.6$
- Affiliation-motivated: $\Omega^j = 1.1$

We choose these values as the theoretical analysis in the previous section indicates that agents with these OMIs have the most pronounced differences in perception. They are thus interesting to study.

In Figure 5a all agents have a "high" OMI, representing power-motivated individuals of the type that perceive a PD game. We see that, because the perceived games are PD games, all agent pairs tend to converge on the (B^D, B^D) NE (bottom left corner) over time. This outcome is the non-cooperative (*i.e.*, competitive) outcome. Competitive behavior is one of the characteristics of power-motivated individuals recognized by psychologists that we saw in Table 1.

Now suppose A^2 agents are achievement-motivated (prefer moderate incentive). We see from Figure 5b that the outcome is different in this case. A^1 power-motivated agents still prefer competitive behavior (B^D). A^2 agents initially choose to increase or maintain their levels of cooperation, indicated by the upward and horizontal trends of their trajectories. However, A^2 agents learn they are being exploited by A^1 and change their strategy to B^D. This is indicated by the downward trends of their trajectories in Figure 5b, and eventual convergence in the bottom left corner of the figure.
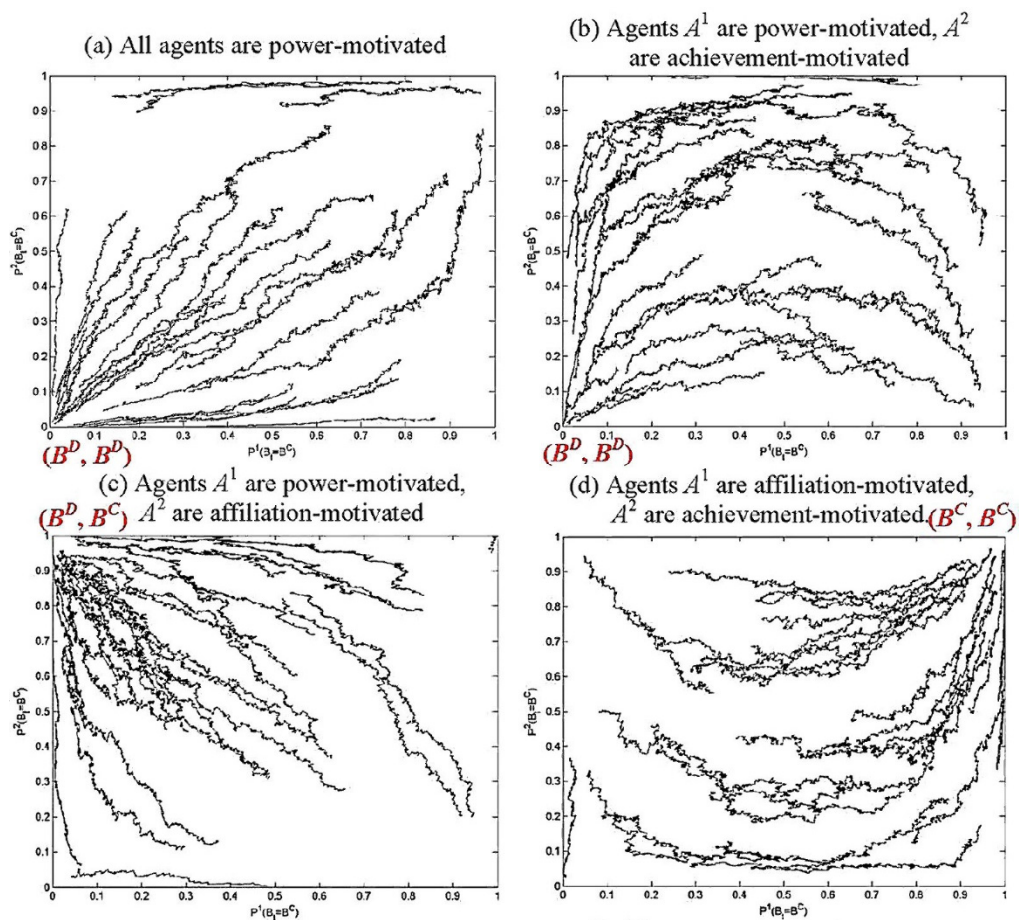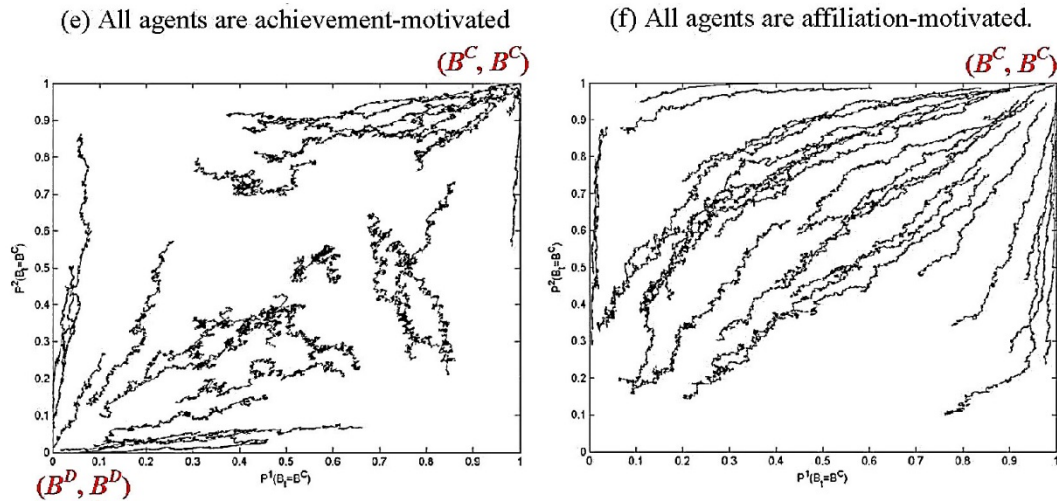

Figure 5. *Cont.*

Figure 5. Learning trajectories when six sets of r thirty pairs of motivated learning agents play 3000 iterations of the prisoners' dilemma game. Each set of agents has differently motivated opposing agents as per the individual sub-figure labels. Trajectories start at random positions in the centre of each sub-figure at *t* = 0, and proceed towards one or more of the corners. Each corner represents a different equilibrium, as per the labels in (**a**). (**a**) All agents are power-motivated; (**b**) Agents A^1 are power-motivated, A^2 are achievement-motivated; (**c**) Agents A^1 are power-motivated, A^2 are affiliation-motivated; (**d**) Agents A^1 are affiliation-motivated, A^2 are achievement-motivated; (**e**) All agents are achievement-motivated; (**f**) All agents are affiliation-motivated.

Next, we consider what happens when agents A^2 are affiliation motivated. We see in Figure 5c that the power-motivated agents A^1 continue to exhibit a preference for competitive behavior B^D. Their trajectories move directly from the right to the left of this figure In contrast, agents A^2 choose the cooperative solution, resulting in convergence in the top left corner of this figure. They do this in spite of being universally exploited by their opponents A^1 as a result.

In contrast to an affiliation-motivated agent, an achievement-motivated agent will learn to adapt to the strategy of its opponent. We have already seen that an achievement-motivated agent will eventually learn to choose B^D if its opponent attempts to exploit it. Figure 5d shows that an achievement-motivated agent will, however, choose B^C, when it plays an opponent that also prefers B^C. Agents A^2 are achievement-motivated, while agents A^1 are affiliation-motivated in Figure 5d. The U shaped learning trajectories shows that achievement-motivated agents A^2 will respond to an opponent choosing B^D by also choosing B^D, resulting in the initial downward trend of the learning trajectory. However, when the opponent begins to cooperate, the achievement-motivated agent also begins to cooperate, leading to the upward moving trajectory and convergence in the top right corner of the figure. In summary, achievement-motivated agents in will adopt either an aggressive, competitive strategy or a cooperative strategy, depending on the behavior exhibited by their opponent. When two achievement-motivated agents play each other, they will converge on either the (B^D, B^D) or (B^C, B^C) outcome, depending on the initial choice of both agents (Figure 5e).

When both players have $\Omega^j < 1.5$, the (B^C, B^C) outcome is dominant as shown in Figure 5f and predicted by Theorem 6. This is, of course, the cooperative outcome and cooperative or collaborative

behavior is one of the recognized characteristics of affiliation-motivated individuals that we saw in Table 1.

We note that with a learning rate of α = 0.001 it takes a few thousand iterations before behavior is observed that converges on the game's theoretical equilibrium. This gradual adaptation can be advantageous for agents who require lifelong learning. Alternatively, speed of learning can be increased by increasing α. The effect of increasing α is to permit the agent to make larger adjustments to is values of $P^j(B_0 = B^C)$ and $P^j(B_0 = B^D)$. This effect is illustrated in Figure 6 for α = 0.01 and α = 0.1. Each tenfold increase in α also speeds convergence roughly tenfold, meaning that ten times fewer decisions are made before $P^1(B_0 = B^C)$ and $P^2(B_0 = B^C)$ stabilize. The trade-off is that, particularly in the case of α = 0.1, the learning trajectory is less predictable. The agents can have large changes in $P^j(B_0 = B^C)$ and these changes can be either increases or decreases. In summary, while the eventual equilibrium remains predictable, the trajectory taken to get there can be highly variable.
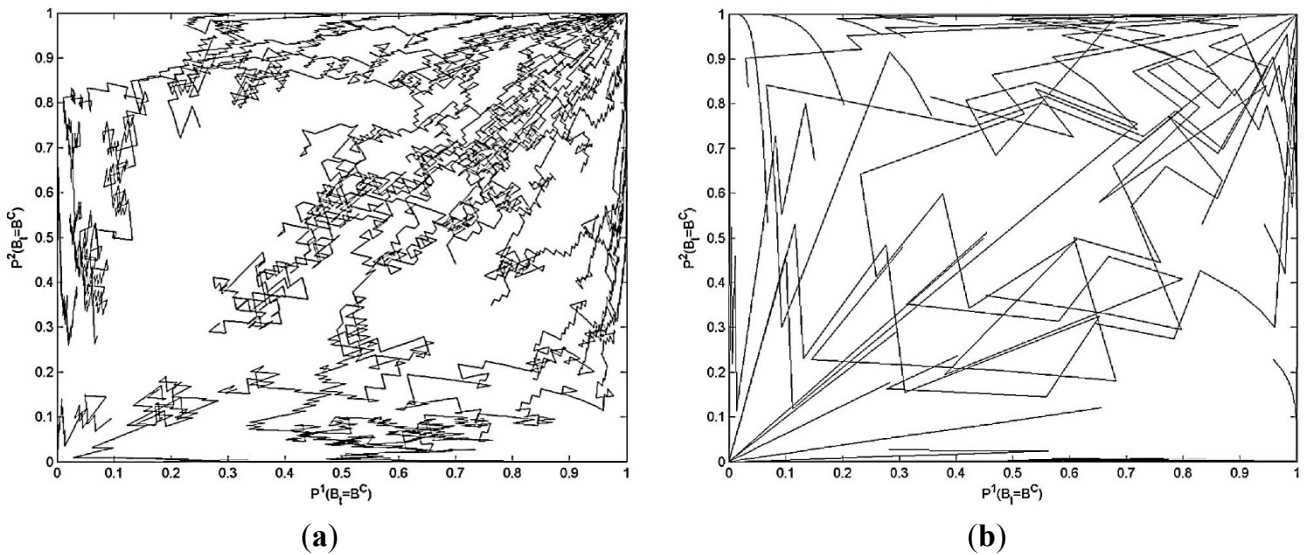

(a)   **(b)**

Figure 6. Thirty pairs of motivated learning agents play 3000 iterations of the prisoner's dilemma game. All agents are affiliation-motivated. Agents in (**a**) have α = 0.01. Agents in (**b**) have α = 0.1.

From the experiments above, in scenarios that can be modeled by a PD game we can conclude that:

- Power-motivated agents, specifically those with $\Omega^j > \frac{1}{2}(T + R)(n - 1)$, will adapt to exploit (choose B^D) all opponents. They will exhibit characteristics of competitive behavior.
- Achievement-motivated agents, specifically those with $\frac{1}{2}(T + P)(n - 1) > \Omega^j > \frac{1}{2}(R + P)(n - 1)$ will adapt differently to different opponents, choosing B^D when exploited, but B^C when their opponent does likewise.
- Affiliation-motivated agents, specifically those with $\Omega^j < \frac{1}{2}(P + S)(n - 1)$, will choose B^C against all opponents. They will exhibit characteristics of cooperative behavior.

These simulations give us an initial understanding of the adaptive diversity that can be achieved using motivated learning agents. There are clear differences in perception by agent with different motivations, and these differences manifest in behavior when they interact. There is some correlation

between the characteristics that psychologists associate with different motivation types and the types of behavior exhibited by agents with different motives.

3.1.3. Empirical Study of the Evolution of Motivated Agents during *n*-Player Common Pool Resource Games

This section presents an empirical study of the evolution of seven different types of agent playing a CPR game. In contrast to the previous section, which considered how individual agents learn, the *n*-player CPR game permits us to consider how the proportions of agents with different motives change in a society under the constrained conditions of a CPR game. For the CPR game in this section, we use values *T = 4, R = 3, S = 2* and *P = 1*. Specific OMI values for the different types of agents are shown in Table 2.

The first type of agent ($j = 1$) uses the explicit payoff of the game to compute evolutionary fitness. We call these agents "correct perceivers" (CPs). The lattfer six types of agents fall into each of the six categories examined theoretically in Section 3.1.1. They use the game transformations presented in Section 2 to compute $\hat{P}^k$, $\hat{T}^k$, $\hat{S}^k$ and $\hat{R}^k$ which are then compounded to form an *n*-player game in a 14×14 matrix of the form in Equation (4).

Table 2. Experimental setup for studying evolution in the common pool resource game.

j	Name	Description	OMI	$F^j(B_0 = B^C)$	$F^j(B_0 = B^D)$
1	CP	Correct perceiver	n/a	0.5	0.5
2	nPow(1)	Strong power motivated agent	375	0	0
3	nPow(2)	Weak power motivated agent	325	0	0
4	nAch(1)	Achievement motivated agent	275	0	0
5	nAch(2)	Achievement motivated agent	225	0	0
6	nAff(1)	Weak affiliation motivated agent	175	0	0
7	nAff(2)	Strong affiliation motivated agent	125	0	0

We then construct a vector **x**₀ of the form in Equation (9), stipulating the initial proportions of each type of agent choosing B^C and B^D (see the last two columns of Table 2 for values used). Initially the society comprises only CP agents (without motivation) with equal probability of choosing either B^C or B^D. The simulation examines whether there is evolutionary benefit of motivated agents by permitting mutation of motivated agents and allowing survival of agents with the highest fitness. For motivated agents, subjective incentive is used to measure fitness. For non-motivated agents explicit payoff is used to measure fitness.

In the simulations *n* = 101 and *h* = 0.001. Each simulation is run for 3000 generations. Mutation is governed by the matrix **Q** shown in Equation (19). Each type of agent has a probability of 0.98 of no mutations from within that type. Correct perceivers have a probability of 0.02 of mutating to nPow(1) agents, who themselves have an equal probability of choosing B^C or B^D. All the motivated agents have a probability of 0.02 of mutating to another type of agent with a slightly lower or slightly higher OMI.

$$
\mathbf{Q} = \begin{bmatrix}
0.980 & 0.000 & 0.005 & 0.005 & \cdots & 0.000 & 0.000 \\
0.000 & 0.980 & 0.005 & 0.005 & \cdots & 0.000 & 0.000 \\
0.010 & 0.010 & 0.980 & 0.000 & & 0.000 & 0.000 \\
0.010 & 0.010 & 0.000 & 0.980 & \cdots & 0.000 & 0.000 \\
0.000 & 0.000 & 0.005 & 0.005 & & 0.000 & 0.000 \\
0.000 & 0.000 & 0.005 & 0.005 & \cdots & 0.000 & 0.000 \\
 & & & \cdot & & & \\
 & & & \cdot & & & \\
 & & & \cdot & & & \\
0.000 & 0.000 & 0.000 & 0.000 & & 0.010 & 0.010 \\
0.000 & 0.000 & 0.000 & 0.000 & \cdots & 0.010 & 0.010 \\
0.000 & 0.000 & 0.000 & 0.000 & & 0.980 & 0.000 \\
0.000 & 0.000 & 0.000 & 0.000 & \cdots & 0.000 & 0.980
\end{bmatrix}
\tag{19}
$$

We examine three charts that expose the population dynamics that occur when different motivated and non-motivated agents (CPs) interact in a multi-player game. The first chart shows the fraction of each type of agent in the population in each generation. That is, $\sum_i F^j(B_t = B^i)$. The second chart shows the probability with which agents of given type will choose B^C at the end of the 3000th generation, as defined in Equation (10). Finally, the third chart shows the subjective incentive perceived by each of the motivated agent types.

Figure 7 shows that the proportion of CPs drops dramatically in the first 100 generations as mutations progressively introduce different kinds of motivated agents into the population. We see that all of these mutants survive and thrive from generation to generation, but some form a greater on-going proportion of the population than others. Specifically, nAch(2) agents and nAff(2) agents form approximately 65% of the population by the end of the 3000th generation. Figure 8 shows that the nAch(2) agents prefer the B^D choice, while the nAff(2) agents prefer the B^C choice. In fact, by the 3000th generation 49% of agents prefer the B^C choice and 51% prefer the B^D choice.
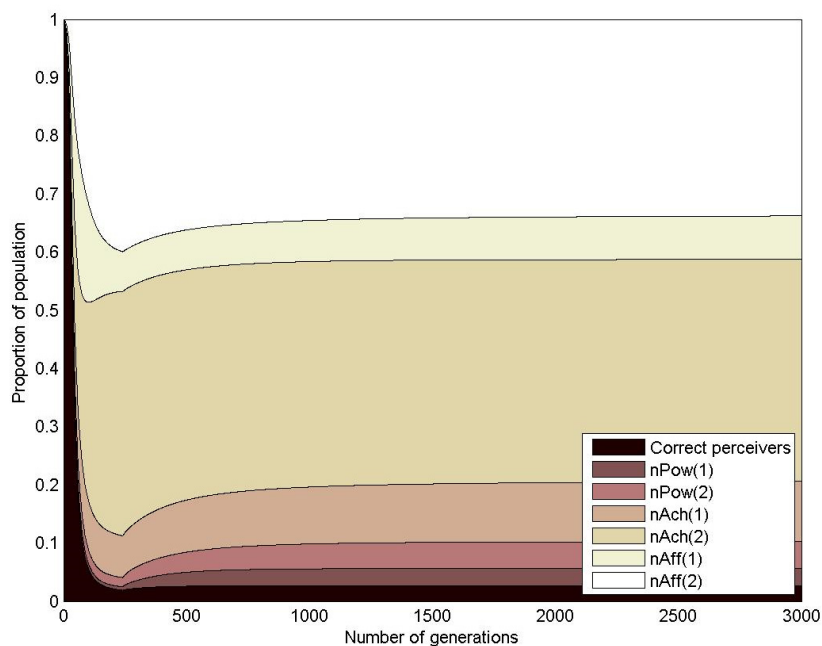

Figure 7. Change in composition of a society of agents playing a common pool resource game.
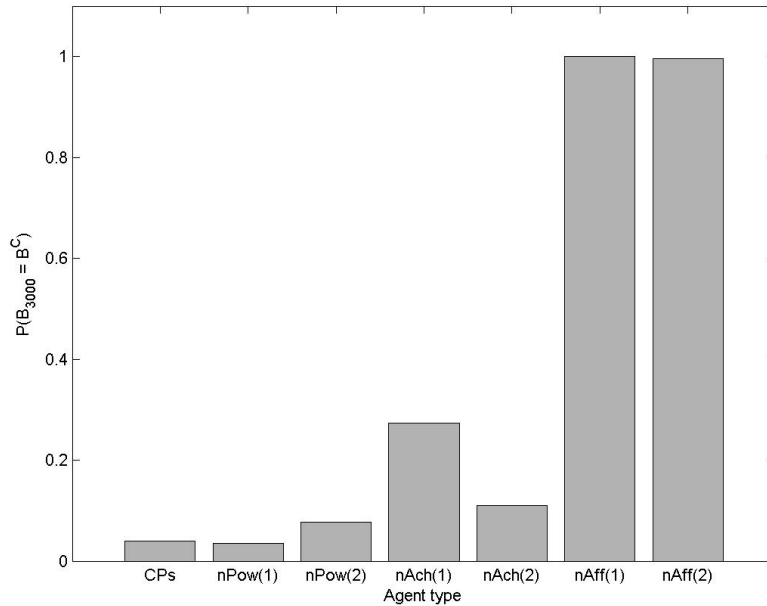
Figure 8. Probability with which each type of agent chooses B^C by the end of the 3000th generation of a common pool resource game.

The NE for a society of CPs predicts that 100% of agents will prefer the B^D choice. The average explicit incentive for all agents would be 200. However, in our simulation with motivated agents, the average explicit payoff is higher: 248.6 on average. This is because some agents are subjectively fitter choosing B^C (see Figure 9). This cooperation raises the objective fitness of the population. In this experiment we thus see an evolutionary benefit of motivation: specifically that differences in subjective fitness result in a diversity of agents and higher overall objective fitness of the society.
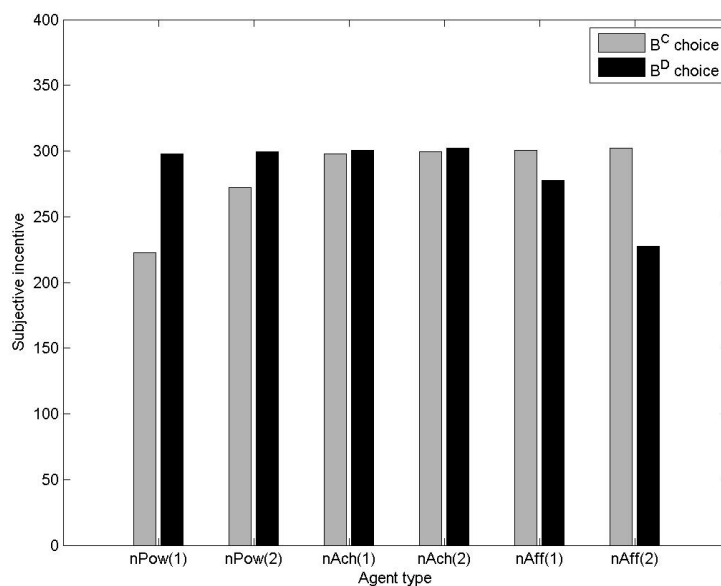

Figure 9. Subjective incentive of each behavior for different types of motivated agent playing a common pool resource game.

The next section examines the theoretical and empirical results for a second game, and provides evidence that the properties we observe in this section hold in other scenarios.

3.2. Snowdrift and the Hawk-Dove Game

The snowdrift game occurs when two drivers are stuck at a snowdrift. Each driver has the option of shoveling snow to clear a path (B^C), or remaining in their car (B^D). The highest payoff outcome T is to leave the opponent to clear all the snow. The opponent receives a payoff of S in this case. However, if neither player clears the snow, then neither can traverse the drift and both receive the lowest possible payoff of P. A reward of R to both results if both players clear the snow. A snowdrift games occurs when $T > R > S > P$.

The *n*-player compound of a snowdrift game is the hawk-dove game [39]. Traditionally, the hawk-dove game is a contest over resources such as food or a mate. The contestants in the game are labeled as either "hawks" or "doves"[1]. The strategy of the "hawk" is to first display aggression, then escalate into a fight until it either wins or is injured. The strategy of the "dove" is to first display aggression, but run to safety if the opponent escalates to a fight. If not faced with this level of escalation the dove will attempt to share the resource.

The contested resource in a hawk-dove game is given the value U, and the damage from losing a fight is given a cost C. C is assumed to be greater than U. Thus we have the following possibilities during pairwise interactions:

- A hawk meets a dove and the hawk gets the full resource. Thus $T = U$.
- A hawk meets another hawk of equal strength. Each wins half the time and loses half the time. Their average payoff is thus $P = \frac{U}{2} - \frac{C}{2}$ each. Note that P is negative.
- A dove meets a hawk. The dove backs off and gets nothing (that is, $S = 0$)
- A dove meets a dove and both share the resource ($R = \frac{U}{2}$ each).

As discussed previously, these payoffs are compounded to form an *n*-player game. That is, $\frac{(n-1)U}{2}$ is the reward if all players choose B^C, $\frac{(n-1)(U-C)}{2}$ is the punishment if all players choose B^D, $(n-1)U$ is the temptation to choose B^D when all other players choose B^C and 0 is the payoff for choosing B^C when the other players choose B^D.

3.2.1. Theoretical Evaluation

We can follow the same process as Section 3.1.1 to construct the perceived versions of snowdrift/the hawk-dove game. For the snowdrift/hawk-dove game, agent types are defined by the following OMI ranges:

- Power-motivated agents have $T(n-1) > \Omega^j > \frac{1}{2}(T+S)(n-1)$
- Achievement-motivated agents have $\frac{1}{2}(T+S)(n-1) > \Omega^j > \frac{1}{2}(R+P)(n-1)$
- Affiliation-motivated agents have $\frac{1}{2}(R+P)(n-1) > \Omega^j > P(n-1)$

Proofs are omitted from section. They follow a similar logic to that used in Appendix A.
Power-Motivated Perception

---

[1] While in nature these are two different species that cannot cross-breed, the spirit of this game model is such that the terms "hawk" and "dove" refer to strategies used in the contest rather than species of bird.

For power-motivated agents playing a snowdrift/hawk-dove game, we assume $T(n - 1) > \Omega^j > \frac{1}{2}(T + S)(n - 1)$. This gives the transformation in Equations (13)–(16). There are two possible perceived games, depending on whether Ω^j is closer to $T(n - 1)$ or $R(n - 1)$. When Ω^j is closer to $T(n - 1)$ we have:

> *Theorem 7:* When a player with $T(n - 1) > \Omega^j > \frac{1}{2}(T + R)(n - 1)$ perceives a snowdrift/hawk-dove game **W** with $T > R > S > P$, the game they perceive is still a valid snowdrift/hawk-dove game $\hat{T}^j > \hat{R}^j > \hat{S}^j > \hat{P}^j$.

When Ω^j is closer to $R(n - 1)$, the additional transformation in Equation (17) is possible. A single perceived game still results as follows:

> *Theorem 8:* When a player with $\frac{1}{2}(T + R)(n - 1) > \Omega^j > \frac{1}{2}(T + S)(n - 1)$ perceives a snowdrift/hawk-dove game **W** with $T > R > S > P$, the game they perceive will have $\hat{R}^j > \hat{T}^j > \hat{S}^j > \hat{P}^j$.

Figure 10 visualizes the structure of the perceived games in Theorems 7 and 8. The visualization shows that the expected subjective incentive changes as OMI decreases. Power-motivated agents with OMIs in the highest range expect to prefer B^C if most other agents prefer B^D. A mixed strategy equilibrium is also possible. However, power-motivated agents with a lower OMI perceive a game in which the "always play B^C" strategy dominates.
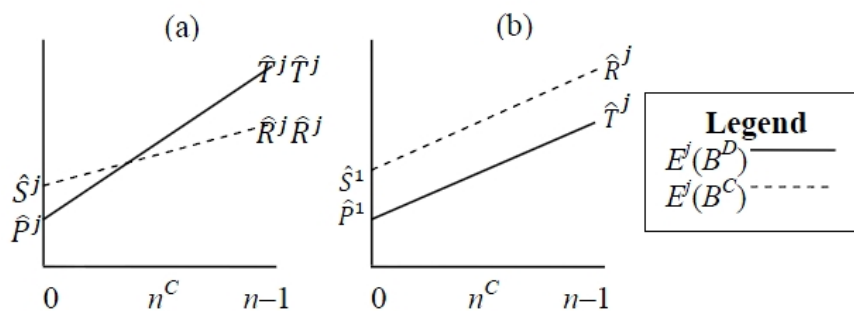

Figure 10. Visualization of the subjective incentive perceived by power-motivated agents playing an explicit hawk-dove game according to (**a**) Theorem 7 and (**b**) Theorem 8.

Achievement-Motivated Perception

For achievement motivated individuals we assume $\frac{1}{2}(T + S)(n - 1) > \Omega^j > \frac{1}{2}(R + P)(n - 1)$. We again consider two cases, depending on whether Ω^j falls closer to $R(n - 1)$ (Theorem 9) or $S(n - 1)$ (Theorem 10). The transformations in Equations (13)–(18) and (20) are all required.

$$\hat{S}^j = T(n - 1) - (S(n - 1) - \Omega^j) = (T - S)(n - 1) + \Omega^j \tag{20}$$

> *Theorem 9:* When a player with $\frac{1}{2}(T + S)(n - 1) > \Omega^j > \frac{1}{2}(R + S)(n - 1)$ perceives a snowdrift/hawk-dove game **W** with $T > R > S > P$, the game they perceive is either
> $\hat{R}^j > \hat{S}^j > \hat{T}^j > \hat{P}^j$ if $\Omega^j > \frac{1}{2}(T + P)(n - 1)$ or
> $\hat{R}^j > \hat{S}^j > \hat{P}^j > \hat{T}^j$ if $\frac{1}{2}(T + P)(n - 1) > \Omega^j$.

Theorem 10: When a player with ½($R + S$)($n − 1$) > Ω^j > ½($R + P$)($n − 1$) perceives a snowdrift/hawk-dove game with $T > R > S > P$, the perceived game is either:

$\hat{S}^j > \hat{R}^j > \hat{T}^j > \hat{P}^j$ if Ω^j > ½($T + P$)($n − 1$) or

$\hat{S}^j > \hat{R}^j > \hat{P}^j > \hat{T}^j$ if ½($T + P$)($n − 1$) > Ω^j

In all four cases that result from Theorems 9 and 10, the strategy "always play B^C" dominates. Achievers will always dig in the snowdrift game, even if it means working alone. Likewise, they will always play a dove strategy in the hawk-dove game.

Affiliation-Motivated Perception

For affiliation-motivated agents we assume ½($R + P$)($n − 1$) > Ω^j > $P(n − 1)$. We consider two cases depending on whether Ω^j is closer to $S(n − 1)$ (Theorem 11) or $P(n − 1)$ (Theorem 12). The additional transformation in Equation (20) is required for the case when $\Omega^j < S(n − 1)$.

Theorem 11: When a player with ½($R + P$)($n − 1$) > Ω^j > ½($P + S$)($n − 1$) perceives a snowdrift/hawk-dove game with $T > R > S > P$, the perceived game is $\hat{S}^j > \hat{P}^j > \hat{R}^j > \hat{T}^j$.

Theorem 12: When a player with ½($S + P$)($n − 1$) > Ω^j > $P(n − 1)$ perceives a snowdrift/hawk-dove game **W** with $T > R > S > P$, the game they perceive is $\hat{P}^j > \hat{S}^j > \hat{R}^j > \hat{T}^j$.

Figure 11 visualizes the structure of the perceived games in Theorems 11 and 12. In the first game (Figure 11a) the "always play B^C" strategy dominates. However, this ceases to be the case for agents with very low OMIs (Figure 11b). These agents prefer to play the same strategy as the majority of other agents is adopting.
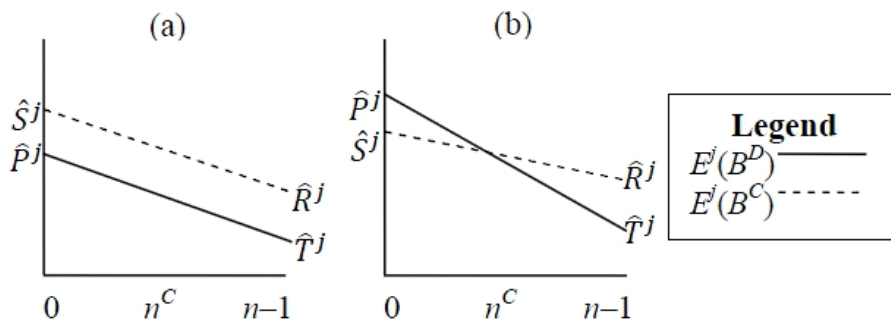

Figure 11. Visualization of the subjective incentive perceived by affiliation-motivated agents playing an explicit hawk-dove game according to (**a**) Theorem 11 and (**b**) Theorem 12.

It is clear from the analysis above that agents with different dominant motives also perceive the snowdrift/hawk-dove game differently. The perceived game depends on the OMI of the agent and the distribution of the explicit payoff. The next section considers how these differences in perception influence learning when agents interact with other agents with different motives.

3.2.2. Learning in the Snowdrift Game

The results in this section are presented in the same order as Section 3.1.2, starting with power-motivated agents. When both players are power-motivated, they both perceive a snowdrift game. Three strategies emerge over time as shown in Figure 12a. Players either converge on the (B^C, B^D) or (B^D, B^C) outcome (top left, or bottom right corners), or their trajectories remain in the central region of the figure. This means they prefer a mixed strategy in which B^D and B^C are chosen with some probability. The equilibrium that emerges depends on the initial probabilities of choosing B^C or B^D.

When one of the players is power-motivated and the other is not, different equilibria emerge. Figure 12b shows power-motivated agents A^1 playing achievement-motivated agents A^2. We see that, when competing with an achievement-motivated agent that has an initial preference for refusing to dig, power-motivated agents will initially increase their probability of digging. However, once the achievement-motivated agent starts to do the same, the power-motivated agents learn to exploit this and refuse to dig. Eventually an equilibrium is reached with power-motivated agents choosing B^D and achievement-motivated agents choosing B^C. This corresponds to the top left of the figure.
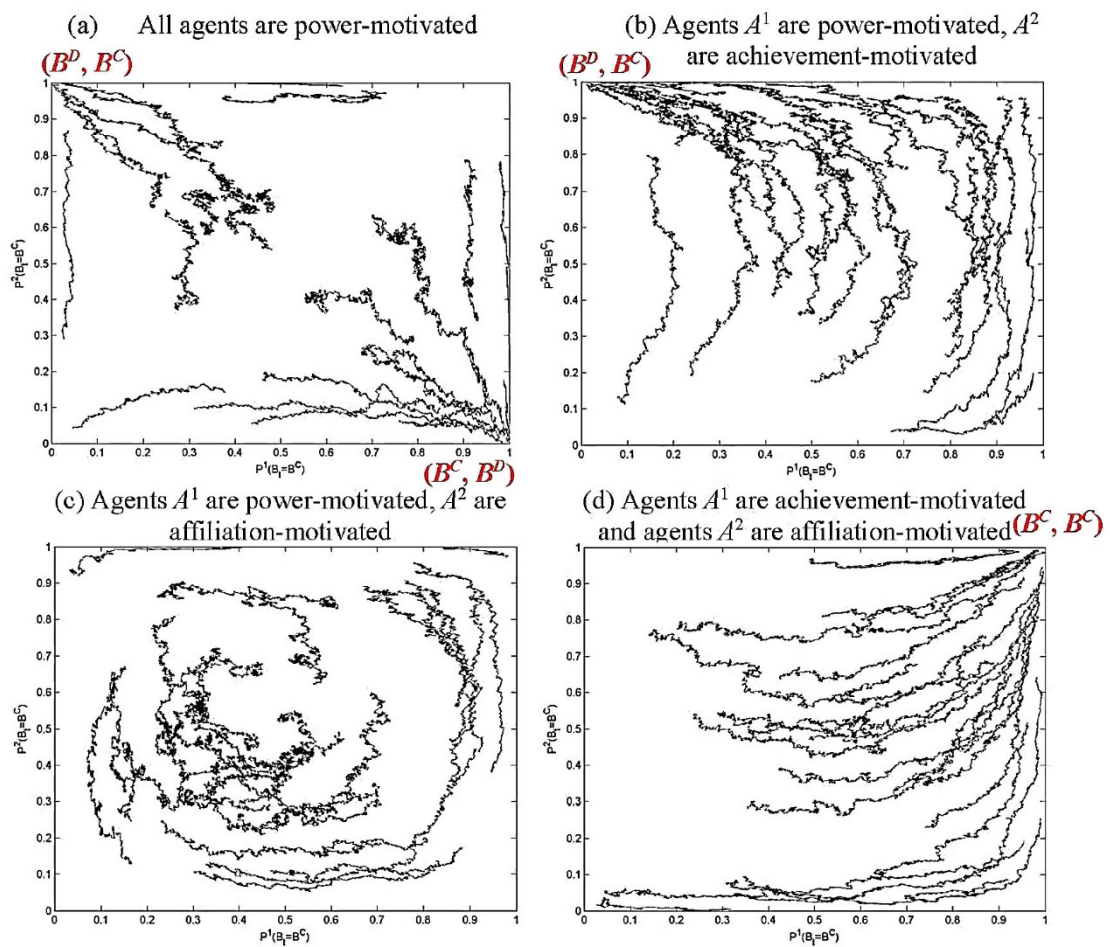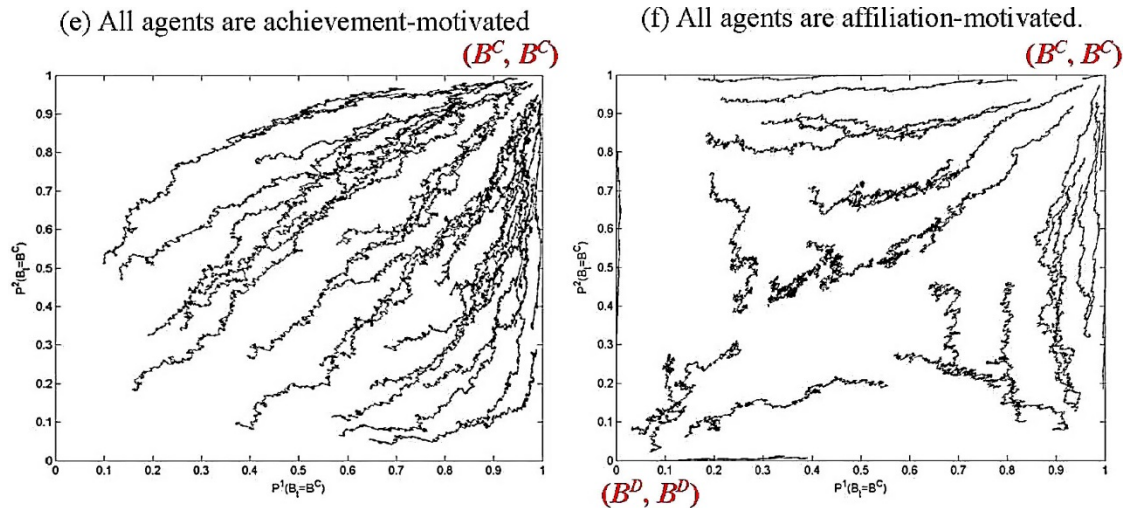

Figure 12. *Cont.*

Figure 12. Learning trajectories when six sets of r thirty pairs of motivated learning agents play 3000 iterations of the snowdrift game. Each set of agents has differently motivated opposing agents as per the individual sub-figure labels. Trajectories start at random positions in the centre of each sub-figure at $t = 0$, and proceed towards one or more of the corners. Each corner represents a different equilibrium, as per the labels in (**a**). (**a**) All agents are power-motivated; (**b**) Agents A^1 are power-motivated, A^2 are achievement-motivated; (**c**) Agents A^1 are power-motivated, A^2 are affiliation-motivated; (**d**) Agents A^1 are achievement-motivated and agents A^2 are affiliation-motivated; (**e**) All agents are achievement-motivated; (**f**) All agents are affiliation-motivated.

Interestingly, this equilibrium is not reached when power-motivated agents encounter affiliation-motivated agents. Affiliation-motivated agents prefer outcomes in which both players do the same thing (see Figure 12f). That is, outcomes when both players cooperate and dig together or both make the decision to refuse to dig. Thus, when the power-motivated agents begin to exploit affiliation-motivated agents, affiliation-motivated agents will learn to refuse to dig. Thus Figure 12c shows the emergence of an ongoing cycle emerges between four pure strategy equilibria.

This oscillation disappears when the difference in OMIs decreases and is replaced by cooperation. This cooperation is evident between almost any pairs of achievement and affiliation motivated agents (see Figure 12d–f). However, because affiliation motivated agents are subjectively content if they are doing the same as their opponent, some pairs of affiliation-motivated agents will also mutually refuse to dig (Figure 12f).

In summary, in scenarios that can be modeled by a snowdrift game, we can conclude that:

- Power-motivated agents, specifically those with $\Omega^j > \frac{1}{2}(T + R)(n - 1)$, prefer outcomes in which one player refuses to dig (either (B^C, B^D) or (B^D, B^C)). If one player is not power-motivated then the power-motivated agent will be the one that refuses to dig. That is, the power-motivated agent will gain the benefit from the digging without exerting themselves. This is consistent with the preference for competitive behavior identified in Table 1.
- Achievement-motivated agents, specifically those with $\Omega^j = \frac{1}{2}(R + S)(n - 1)$ will dig regardless of the motive profile of the other player. They do not mind working alone, consistent with the suggestion in Table 1.

- Affiliation-motivated agents, specifically those with $\Omega^j < \frac{1}{2}(P + S)\,(n - 1)$, prefer the outcomes where both players do the same thing. This is consistent with wanting to belong to a (peer) group, as in Table 1.

Adaptive diversity is again demonstrated, with different agent types exhibiting different behavior against different opponents. The adaptive characteristics of the three types of agents playing snowdrift are slightly different to when they play a PD game, as they are automatically tailored to the incentives offered in the current game, but the broad characteristics of motivated behavior still hold. There is some similarity between these behaviors and some of the characteristics listed for humans, albeit in a simplified and abstracted manner.

3.2.3. Empirical Study of Evolution in the Hawk-Dove Game

We now examine evolution in the hawk-dove game. We use the assumption that $U = 4$ and $C = 8$. This gives us the non-compounded payoff values $T = 4$, $R = 2$, $S = 0$ and $P = -2$. The game world **W** is again a 14×14 matrix comprised of subjective payoff functions for seven different types of agents, defined as per Table 3.

Table 3. Experimental setup for studying evolution in the hawk-dove game.

j	Name	Description	OMI	$F^j(B_0 = B^C)$	$F^j(B_0 = B^D)$
1	CP	Correct perceiver	n/a	0.5	0.5
2	nPow(1)	Strong power motivated agent	350	0	0
3	nPow(2)	Weak power motivated agent	250	0	0
4	nAch(1)	Achievement motivated agent	150	0	0
5	nAch(2)	Achievement motivated agent	50	0	0
6	nAff(1)	Weak affiliation motivated agent	−50	0	0
7	nAff(2)	Strong affiliation motivated agent	−150	0	0

Figure 13 shows that the proportion of CPs again drops over the first 100 generations as mutations progressively introduce different kinds of motivated agents into the population. In this scenario, we see that not all of these mutants survive and thrive from generation to generation. Some form significant, stable proportions of the population but others do not increase in number beyond the 1% mutation rate. B^C choosing nAch(1) agents form the majority (almost 80%) of the population by the 500th generation, and remain so by the 3000th generation. CPs and nPow(2) agents are the next most prevalent agent types.

The NE for a society of CPs predicts that 50% of agents will choose B^D and 50% B^C. The average explicit incentive for all agents would be 100. However, in our simulation with motivated agents, the average explicit incentive is higher: 194.1 on average. This is because more than 50% of agents are subjectively fitter choosing B^C (see Figures 14 and 15). This cooperation raises the objective fitness of the population. In this experiment we thus again see an evolutionary benefit of motivation, as well as some population diversity.
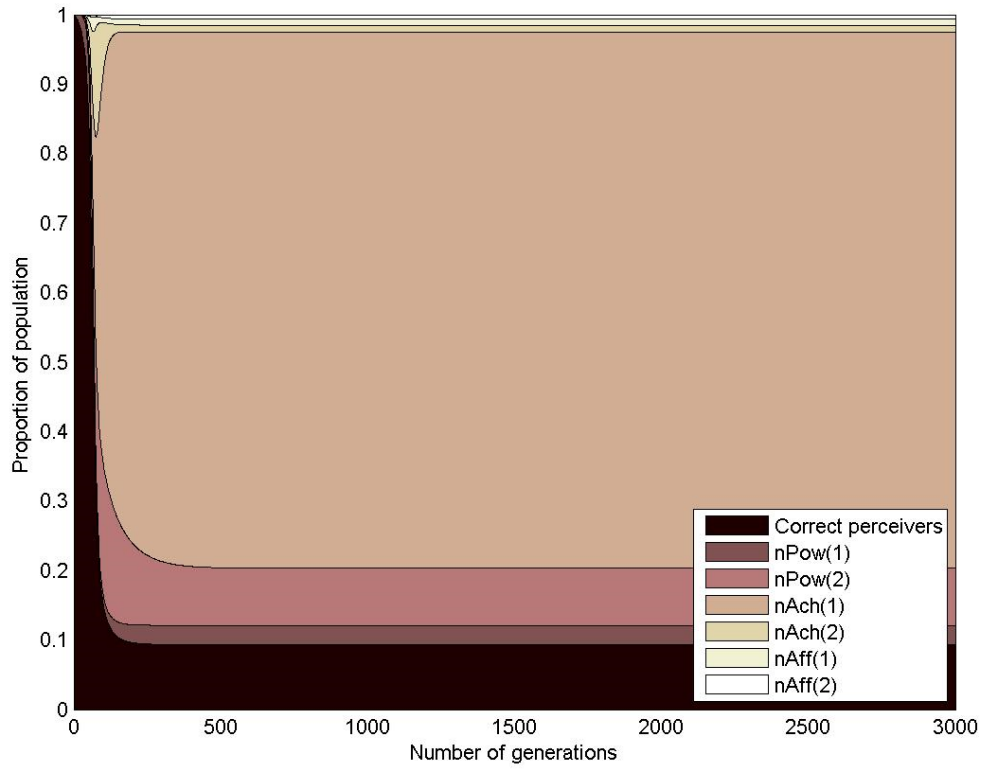
Figure 13. Change in composition of a society of agents playing a hawk-dove game.
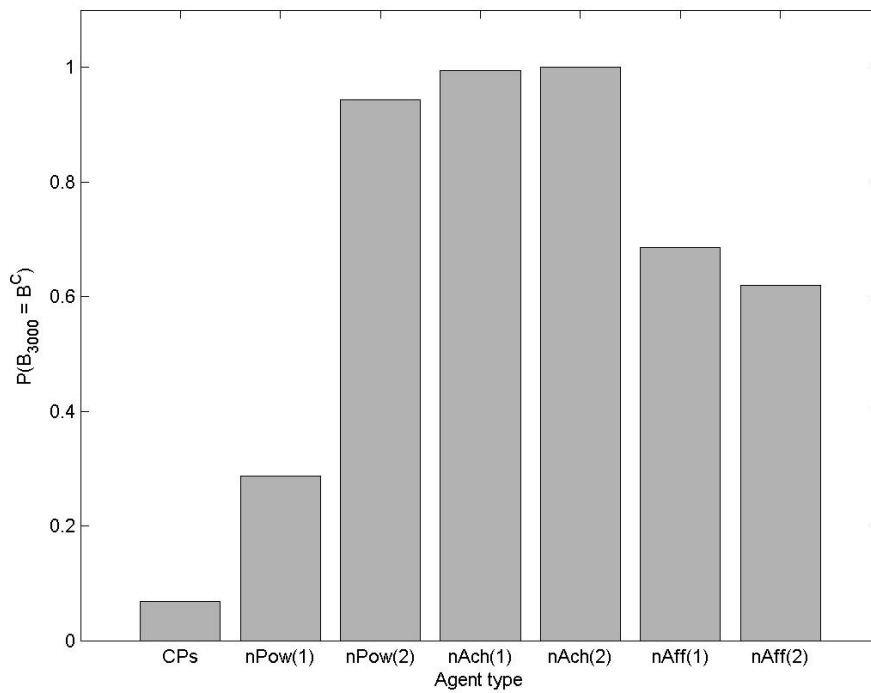

Figure 14. Probability with which each type of agent chooses B^C by the end of the 3000th generation of a hawk-dove game.
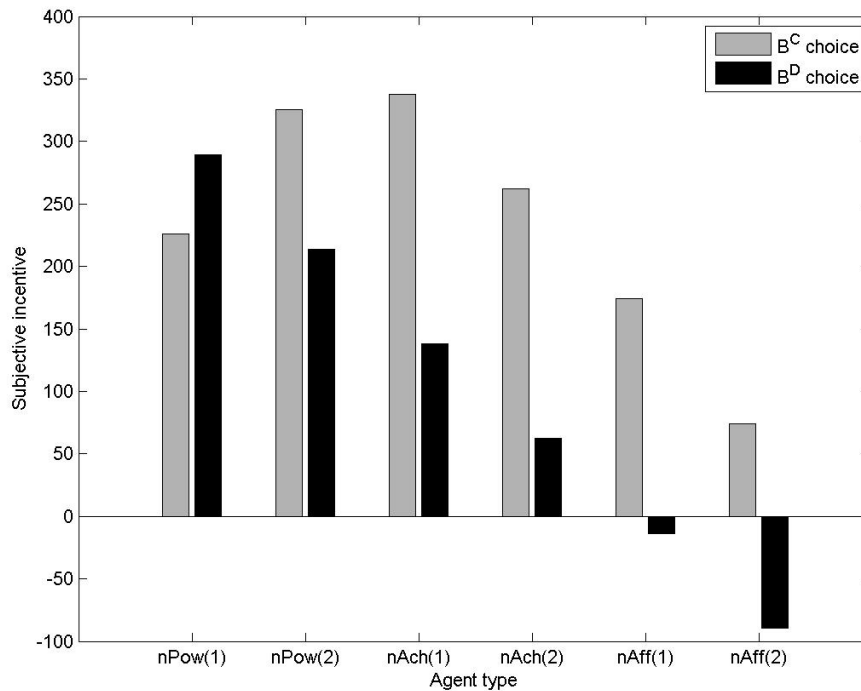
Figure 15. Subjective incentive of each behavior for different types of motivated agent playing a hawk-dove game.

4. Conclusions and Future

In conclusion, the theoretical and experimental analysis in this paper has demonstrated that we can achieve some of the characteristics of particular dominant motives observed in humans in artificial agents. In two two-player games we observed that:

- Power-motivated agents learn competitive behavior.
- Achievement-motivated agents will adapt differently to different opponents and different games.
- Affiliation-motivated agents will exhibit characteristics of cooperative behavior and prefer outcomes where both players do the same thing.

When an evolutionary algorithm is used to create new generations of motivated agents:

- Different types of motivated agents thrive in different scenarios
- Diversity of agents is achieved
- Evolutionary benefit can be observed in the form of higher average explicit incentive (payoff) achieved by the society than we would expect of a society of objectively rational agents without motivation.

These conclusions are observed in both of the games studied, providing evidence that the methodology and results are applicable beyond a single scenario.

From a human perspective, the models explored in this paper give us some insight into to the relationship between incentive, motivation and emergent behavioral characteristics. The computational models support the hypothesis that competitive behavior is related to a preference for higher incentive goals. Conversely, cooperative behavior emerges in agents with a preference for lower incentive goals.

In demonstrating this, the models in this paper better reflect the decision-making diversity seen in humans, than do traditional models that assume objective rationality. .

However, there are other salient characteristics of the various dominant motives listed in Table 1 that are not revealed in the scenarios studied in this paper. Different and potentially more complex abstract game scenarios are likely to be required to reveal these characteristics during game play. Game theory literature provides us with a rich set of alternatives for this. Some examples that lend themselves to study in the context of motivation include the leader game, battle of the sexes, stag hunt and public goods games.

The next phase of this work will continue to examine the behavior of agents when they are applied in more concrete settings. Specifically, we are interested in the emergent behavior of agents when B^C and B^D have an effect on a real or simulated environment. Some scenarios that could be used as starting points have been proposed by Givigi and Schwarz [28], who examined the behavior of agents with different personality traits. A number of variations of such scenarios are possible, including the interaction of two motivated agents, the interaction of a society of motivated agents, or the interaction of a single motivated agent with many other (not necessarily motivated) agents, or even humans. Other settings for this kind of investigation include computer games and virtual worlds, where humans are used to interacting with virtual agents. Motivated agents offer a new kind of behavioral diversity in this setting, while giving game programmers the confidence of the theoretical predictions governing the emergent learning behavior of agents.

The study of computational motivation in a game theoretic setting offers us a way to model and understand diversity in decision-making by acknowledging individual differences in implicit motives. Implementation of computational motivation in artificial agents permits us to simulate strategic interactions at both an individual and social level. Because we can run large numbers of simulations rapidly, there is, in future, potential to explore "what-if" scenarios in simulation. Such scenarios may include exploration of the impact of changing payoff functions on decision-making by individuals or groups. This may further permit development of rules that explain differences in decision-making as a result of motivation, and suggest how individuals with different motives will respond to changing payoff functions over time.

Conflicts of Interest

The authors declare no conflict of interest.

Appendix A—Proofs

Proof of Theorem 1: If we assume $\hat{R}^j \geq \hat{T}^j$ then we have $(T + R)(n - 1) - \Omega^j \geq \Omega^j$ which simplifies to $\frac{1}{2}(T + R)(n - 1) \geq \Omega^j$. This contradicts the assumption that $T(n - 1) > \Omega^j > \frac{1}{2}(T + R)(n - 1)$ so it must be true that $\hat{T}^j > \hat{R}^j$. If we assume that $\hat{P}^j \geq \hat{R}^j$ then we have $(T + P)(n - 1) - \Omega^j \geq (T + R)(n - 1) - \Omega^j$ or $P \geq R$ which contradicts the definition of PD. Thus it must be true that $\hat{R}^j > \hat{P}^j$. Likewise, if we assume that $\hat{S}^j \geq \hat{P}^j$ then we have $(T + S)(n - 1) - \Omega^j \geq (T + P)(n - 1) - \Omega^j$ which simplifies to $S \geq P$ which contradicts the definition of PD. Thus it must be true that $\hat{P}^j > \hat{S}^j$ $\square$

Proof of Theorem 2: If we assume $\hat{T}^j \geq \hat{R}^j$, there are two possibilities. First, substituting Equations (13) and (14) gives $\Omega^j \geq (T + R)(n - 1) - \Omega^j$ which simplifies to $\Omega^j \geq \frac{1}{2}(T + R)(n - 1)$. This contradicts the assumption that $\frac{1}{2}(T + R)(n - 1) > \Omega^j$. Similarly, substitution of Equations (13) and (17) gives $\Omega^j \geq (T - R)(n - 1) + \Omega^j$ which simplifies to $R \geq T$ and contradicts the definition of the PD. Thus it must be true that $\hat{R}^j > \hat{T}^j$. If we assume that $\hat{S}^j \geq \hat{P}^j$ then we have $(T + S)(n - 1) - \Omega^j \geq (T + P)(n - 1) - \Omega^j$ which simplifies to $S \geq P$. This contradicts the definition of the PD. Thus it must be true that $\hat{P}^j > \hat{S}^j$. Finally, if we assume $\hat{P}^j > \hat{T}^j$ we have $(T + P)(n - 1) - \Omega^j > \Omega^j$ or $\frac{1}{2}(T + P)(n - 1) > \Omega^j$ which contradicts the assumption for power-motivated agents that $\Omega^j > \frac{1}{2}(T + P)(n - 1)$. Thus it must be true that $\hat{T}^j > \hat{P}^j$ □

Proof of Theorem 3: $\hat{R}^j > \hat{T}^j$ and $\hat{P}^j > \hat{S}^j$ according to the proof of Theorem 2. However, as we now assume $\frac{1}{2}(T + P)(n - 1) > \Omega^j$ we have $\hat{P}^j > \hat{T}^j$ (converse of Theorem 2). There are then two alternative orderings, differing in the position of $\hat{S}^j$. If $\hat{T}^j > \hat{S}^j$ substitution of Equations (13) and (16) gives $\Omega^j > (T + S)(n - 1) - \Omega^j$ which simplifies to $\Omega^j > \frac{1}{2}(T + S)(n - 1)$. Conversely if $\hat{S}^j > \hat{T}^j$ then $\frac{1}{2}(T + S)(n - 1) > \Omega^j$ □

Proof of Theorem 4: $\hat{R}^j > \hat{T}^j$ according to the proof of Theorem 2. If we assume that $\hat{S}^j \geq \hat{P}^j$ substituting Equations (15) and (16) gives $(T + S)(n - 1) - \Omega^j \geq (T + P)(n - 1) - \Omega^j$. This simplifies to $S \geq P$ which contradicts the definition of the PD. Likewise, substituting Equations (17) and (18) gives $(T + S)(n - 1) - \Omega^j > (T - P)(n - 1) + \Omega^j$. This simplifies to $\frac{1}{2}(S + P)(n - 1) > \Omega^j$, which contradicts the assumption of an achievement-motivated agent. Thus it must be true that $\hat{P}^j > \hat{S}^j$. There are then two alternative orderings, differing in the position of $\hat{S}^j$. If $\hat{T}^j > \hat{S}^j$ substitution of Equations (13) and (16) gives $\Omega^j > (T + S)(n - 1) - \Omega^j$, which simplifies to $\Omega^j > \frac{1}{2}(T + S)(n - 1)$. Conversely, if $\hat{S}^j > \hat{T}^j$ then $\frac{1}{2}(T + S)(n - 1) > \Omega^j$ □

Proof of Theorem 5: If we assume $\hat{T}^j \geq \hat{R}^j$, substitution of Equations (13) and (17) gives us $\Omega^j \geq (T - R)(n - 1) + \Omega^j$ which simplifies to $R \geq T$. This contradicts the definition of the PD so it must be true that $\hat{R}^j > \hat{T}^j$. If we assume that $\hat{S}^j \geq \hat{P}^j$ substituting Equations (15) and (16) gives us $(T + S)(n - 1) - \Omega^j \geq (T + P)(n - 1) - \Omega^j$ which simplifies to $S \geq P$ which contradicts the definition of the PD. Likewise, substituting Equations (16) and (18) gives $(T + S)(n - 1) - \Omega^j \geq (T - P)(n - 1) + \Omega^j$ which simplifies to $\frac{1}{2}(P + S)(n - 1) \geq \Omega^j$ and contradicts the assumption that $\Omega^j > \frac{1}{2}(P + S)(n - 1)$. Thus it must be true that $\hat{P}^j > \hat{S}^j$. If we assume $\hat{R}^j > \hat{S}^j$ substitution of Equations (16) and (17) gives $(T - R)(n - 1) + \Omega^j > (T + S)(n - 1) - \Omega^j$. Simplification gives $\Omega^j > \frac{1}{2}(R + S)(n - 1)$ which contradicts the assumption for affiliation-motivated agents that $\frac{1}{2}(R + S)(n - 1) > \Omega^j$. Thus it must be true that $\hat{S}^j > \hat{R}^j$ □

Proof of Theorem 6: If we assume $\hat{P}^j \geq \hat{S}^j$ then we have $(T - P)(n - 1) + \Omega^j \geq (T + S)(n - 1) - \Omega^j$ which simplifies to $\Omega^j \geq \frac{1}{2}(P + S)(n - 1)$. This contradicts the assumption that $\frac{1}{2}(P + S)(n - 1) > \Omega^j$ Thus it must be true that $\hat{S}^j > \hat{P}^j$. If we assume $\hat{R}^j \geq \hat{P}^j$ then we have $(T - R)(n - 1) + \Omega^j \geq (T - P)(n - 1) + \Omega^j$ which simplifies to $P \geq R$. This contradicts the definition of the PD. Thus it must be true that $\hat{P}^j > \hat{R}^j$. Likewise, if we assume $\hat{T}^j \geq \hat{R}^j$ then we have $\Omega^j \geq (T - R)(n - 1) + \Omega^j$ which simplifies to $R \geq T$. This contradicts the definition of the PD. Thus it must be true that $\hat{R}^j > \hat{T}^j$ □

Appendix B

This appendix presents a brief demonstration to complete the progression from theoretical and empirical analysis to application. The application chosen here is a simple implementation of Algorithm 1 in the snowdrift scenario. This demonstration takes place in a 16×16 meter space as shown in Figure 16. Two agents A^1 (blue; left) and A^2 (green; right) have exited their cars. Their initial positions (the positions of their cars) are $\mathbf{p}_0^1 = (6, 2)$ and $\mathbf{p}_0^2 = (10, 2)$. The snowdrift is located to the north at location $\mathbf{p}_0^3 = (8, 12)$. As in Section 3.2, $\alpha = 0.001$ and each agent makes 3000 decisions. This occurs over a period of 4 seconds using a MATLAB implementation.
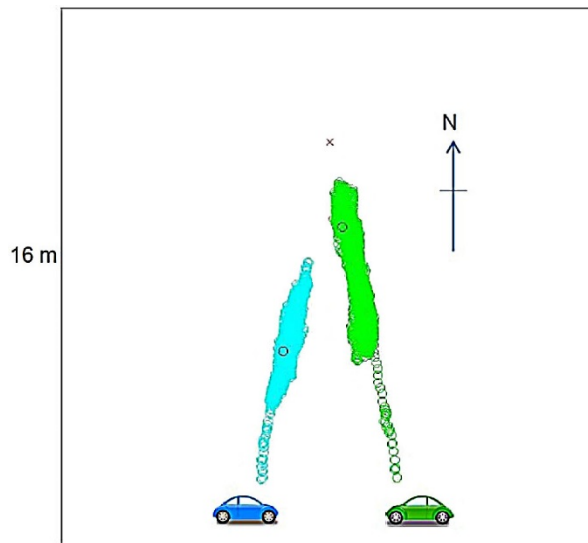

Figure 16. Movement trajectories when two power-motivated agents encounter each other at a snowdrift.

The motivated agents use Algorithm 1 with specific behaviors as follows:

- B^C moves the motivated agent towards the snowdrift according to:

$$\mathbf{p}_{t+1}^j = \mathbf{p}_t^j + \lambda P_t(B_t = B^C)\left(\mathbf{p}_t^3 - \mathbf{p}_t^j\right) + \phi \tag{21}$$

- B^D moves the motivated agent towards position of their own car:

$$\mathbf{p}_{t+1}^j = \mathbf{p}_t^j + \lambda P_t(B_t = B^C)\left(\mathbf{p}_0^j - \mathbf{p}_t^j\right) + \phi \tag{22}$$

where λ is the step-size between 0 and 1 (we used $\lambda = 0.05$) and ϕ is a small random number selected with a uniform distribution from the interval (–0.05, 0.05). This movement update moves the agent approximately in the direction of the snowdrift by a distance that is proportional to the chosen step-size λ and the agent's preference for such movement, represented by $P_t(B_t = B^C)$. Both agents have equal initial preferences for B^C and B^D.

Figure 16 demonstrates two power-motivated agents. These agents never reach the snowdrift, as they both develop mixed strategies, remaining around 50% probability of approaching the drift. In contrast, in Figure 17a, in which a power-motivated and achievement-motivated agent encounter the drift, the achievement-motivated agent approaches the drift, while the power-motivated agent returns to its car. When two achievement-motivated agents encounter the drift they will cooperate to clear it as

shown in Figure 17b. These behaviors conform to the theory and empirical results studied in the previous sections. We can thus achieve a diversity of learned behavior that remains predictable.
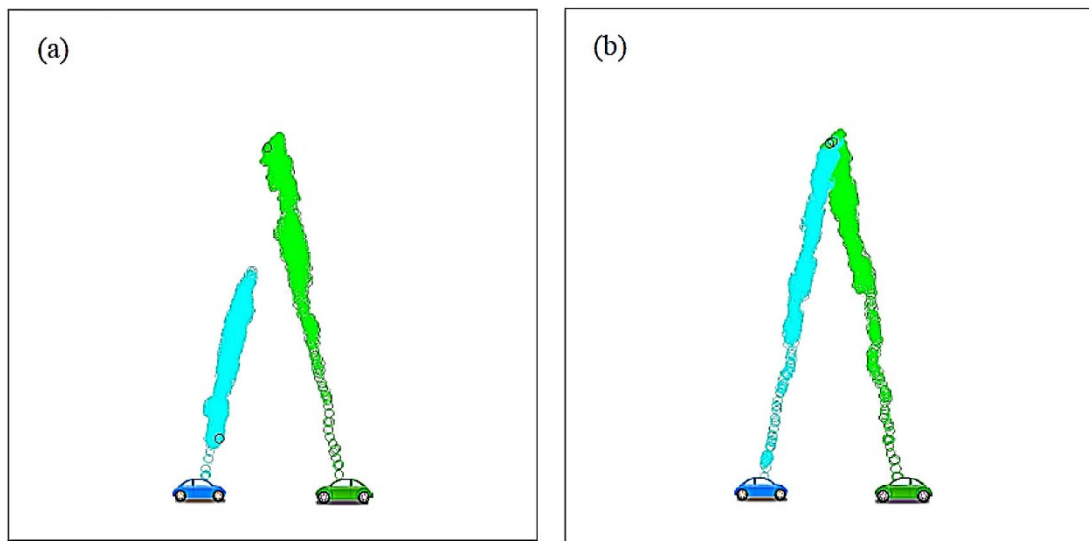

Figure 17. (**a**) A power-motivated (blue) and achievement-motivated (green) agent encounter each other a snowdrift (**b**) Two achievement-motivated agents encounter each other at a snowdrift.

References

1. Heckhausen, J.; Heckhausen, H. *Motivation and Action*; Cambridge University Press: New York, NY, USA, 2010.
2. Terhune, K.W. Motives, situation and interpersonal conflict within prisoner's dilemma. *J. Personal. Soc. Psychol. Monogr. Suppl.* **1968**, *8*, 1–24.
3. Kuhlman, D.; Marshello, A. Individual differences in game motivation as moderators of preprogrammed strategy effects in prisoner's dilemma. *J. Personal. Soc. Psychol.* **1975**, *32*, 922–931.
4. Kuhlman, D.; Wimberley, D. Expectations of choice behavior held by cooperators, competitors and individualists across four classes of experimental game. *J. Personal. Soc. Psychol.* **1976**, *34*, 69–81.
5. Van Run, G.; Liebrand, W. The effects of social motives on behavior in social dilemmas in two cultures. *J. Exp. Soc. Psychol.* **1985**, *21*, 86–102.
6. Atkinson, J.W.; Litwin, G.H. Achievement motive and test anxiety conceived as motive to approach success and motive to avoid failure. *J. Abnorm. Soc. Psychol.* **1960**, *60*, 52–63.
7. Merrick, K.; Maher, M.L. Motivated Reinforcement Learning: Curious Characters for Multiuser Games; Springer: Berlin, Germany, 2009.
8. Merrick, K.; Shafi, K. A game theoretic framework for incentive-based models of intrinsic motivation in artificial systems. *Front. Cogn. Sci. Spec. Issue Intrinsic Motiv. Open-End. Dev. Anim. Hum. Robot.* **2013**, *4*, 1–17.

9. Nguyen, M.; Oudeyer, P.-Y. Socially guided intrinsic motivation for robot learning of motor skills. *Auton. Robot.* **2014**, *36*, 273–394.

10. Baldassare, G.; Mannella, F.; Fiore, V.; Redgrave, P.; Gurney, K.; Mirolli, M. Intrinsically motivated action-outcome learning and goal-based action recall: A system-level bio-constrained computational model. *Neural Netw.* **2013**, *41*, 168–187.

11. Baldassarre, G.; Mirolli, M. Intrinsically Motivated Learning in Natural and Artificial Systems; Springer: Berlin, Heidelberg, Germany, 2013.

12. Oudeyer, P.-Y.; Kaplan, F. *Intelligent Adaptive Curiosity: A Source of Self-Development*; Fourth International Workshop on Epigenetic Robotics, Lund University: Lund, Sweden, 2004; pp. 127–130.

13. Merrick, K.; Shafi, K. Achievement, affiliation and power: Motive profiles for artificial agents. *Adapt. Behav.* **2011**, *19*, 40–62.

14. McClelland, D. *The Achieving Society*; The Free Press: New York, NY, USA, 2010.

15. Sirota, D.; Mischkind, L.; Meltzer, M. *The Enthusiastic Employee*; Pearson Education Inc: Upper Saddle River, NJ, USA, 2005.

16. Atkinson, J.W. Motivational determinants of risk-taking behavior. *Psychol. Rev.* **1957**, *64*, 359–372.

17. Elliot, A.; Eder, A.; Harmon-Jones, E. Approach-avoidance motivation and emotion: Convergence and divergence. *Emot. Rev.* **2013**, *5*, 308–311.

18. Atkinson, J.W.; Raynor, J.O. *Motivation and Achievement*; V.H. Winston: Washington, DC, USA, 1974.

19. Nikitin, J.; Freund, A. When wanting and fearing go together: The effect of co-occurring social approach and avoidance motivation on behavior, affect and cognition. *Eur. J. Soc. Psychol.* **2009**, *40*, 783–804.

20. Elliot, A. *Handbook of Approach and Avoidance Motivation*; Taylor and Francis: New York, NY, USA, 2008.

21. McClelland, J.; Watson, R.I. Power motivation and risk-taking behaviour. *J. Personal.* **1973**, *41*, 121–139.

22. McClelland, J.; Boyatzis, R.E. The leadership motive pattern and long term success in management. *J. Appl. Psychol.* **1982**, *67*, 737–743.

23. Merrick, K. Evolution of intrinsic motives in a multi-player common pool resource game. In Proceedings of the IEEE Symposium Series on Computational Intelligence for Human-like Intelligence, Orlando, FL, USA, 2014; pp. 36–43.

24. Acemoglu, D.; Yildiz, M. *Evolution of Perceptions and Play*; Massachusetts Institute of Technology, Department of Economics: Cambridge, MA, USA, 2001.

25. Dekel, E.; Ely, J.; Ylankaya, O. Evolution of preferences. *Rev. Econ. Stud.* **2007**, *74*, 685–704.

26. Colman, A. Game theory and experimental games: The study of strategic interaction. In *International Series in Experimental Social Psychology*; Pergamon Press: Oxford, UK,1982.

27. Wang, M.; Hipel, K.; Fraser, N. Modeling misperceptions in games. *Behav. Sci.* **1988**, *33*, 207–223.

28. Givigi, S.N.; Schwartz, H.M. Swarm robot systems based on the evolution of personality traits. *Turk. J. Electr.Eng.* **2007**, *15*, 257–282.

29. Nowak, M.; Sigmund, K. Evolution of indirect reciprocity. *Nature* **2005**, *437*, 1291–1298.

30. Wang, Z.; Kokubo, S.; Jusup, M.; Tanimoto, J. Universal scaling for the dilemma strength in evolutionary games. *Phys. Life Rev.* **2015**, *14*, 1–30.

31. Bennett, E. The aspiration approach to predicting coalition formation and payoff distribution in sidepayment games. *Int. J. Game Theory* **1983**, *12*, 1–28.

32. Brumley, L. *Misperception and Its Evolutionary Value*; Monash University: Melbourne, Australia, 2014.

33. Fudenberg, D.; Levine, D. Learning and evolution: Where to we stand? Learning in games. *Eur. Econ. Rev.* **1998**, *42*, 631–639.

34. Chakraborty, D.; Stone, P. Multiagent learning in the presence of memory-bounded agents. *Auton. Agents Multi-Agent Syst.* **2014**, *28*, 182–213.

35. Borgers, T.; Sarin, R. Learning through reinforcement and replicator dynamics. *J. Econ. Theory* **1997**, *77*, 1–14.

36. Schembri, M.; Mirolli, M.; Baldassarre, G. Evolution and learning in an intrinsically motivated reinforcement learning robot. In *Advances in Artificial Life*; Springer: Berlin, Heidelberg, Germany, 2007; Volume 4648, pp. 294–303.

37. Singh, S.; Lewis, R.; Barto, A.G.; Sorg, J. Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Trans. Auton. Ment. Dev.* **2010**, *2*, 70–82.

38. Rapoport, A.; Chammah, A. *Prisoner's Dilemma, A Study in Conflict and Cooperation*; University of Michigan Press: Ann Arbor, MI, USA, 1965.

39. Maynard-Smith, J.; Price, G.R. The logic of animal conflict. *Nature* **1973**, *246*, 15–18.