Chapter 11

Cooperative Trust Games

Dariusz G. Mikulski

Additional information is available at the end of the chapter

http://dx.doi.org/10.5772/53928

1. Introduction

In certain multi-agent systems, the interactions between agents result in the formation of relationships, which can be leveraged for cooperative or collaborative activities. These relationships generally constrain individual-agent actions, since relationships imply that at least one contract (or mutual agreement) between the agents must exist. There is always some uncertainty as to whether or not either agent can or will satisfy some contract requirement – especially at the creation of a new contract. But in order to maintain the existence of a contract, each agent must overcome this uncertainty and assume that the other will do the same. The mechanism that facilitates this "act of faith" is generally regarded as "trust." In essence, each agent (whether a person or organization) in a relationship mutually trusts that the loss of some control will result in cooperative gains that neither agent could achieve alone.

In general, trust helps agents deal with uncertainty by reducing the complexity of expectations in arbitrary situations involving risk, vulnerability, or interdependence [1]. This is because agents rely on trust whenever they need to gauge something they cannot ever know precisely with reasonable time or effort. The benefits of trustworthy relationships include lower defensive monitoring of others, improved cooperation, improved information sharing, and lower levels of conflict [2]. But the reliance on trust also exposes people to vulnerabilities associated with betrayal, since the motivation for trust – the need to believe that things will behave consistently – exposes individuals to potentially undesirable outcomes. Thus, trust is a concept that must not only be managed, but also justified [3].

Since agents in an arbitrary system are always assumed to have selfish interests, the goal of each agent is to try to find the most fruitful relationships in a pool of potential agents [4]. That said, we cannot assume that agents do not already have pre-existing relationships with other agents. Furthermore, some agents may actually be within strongly-connected sub-system groups known as *coalitions*, where every agent in the coalition has a relationship with every other agent in the coalition. A coalition may contain a mixture of trustworthy

and untrustworthy agents – but as a group, achieves cooperative gains that no sub-coalition could match. Thus, agents may be justified in forming relationships with coalition members who are not ideally trustworthy in order to acquire these cooperative gains as well.

As a simple example to illustrate this concept, consider two geographically-separated agents (who never physically met) who would like to conduct a financial transaction in exchange for some good. One agent must provide the good (through the mail) and the other must provide the payment (through the mail or electronically). If both agents follow their economic best interest, then neither agent should participate in the transaction since both agents are vulnerable to betrayal. This is because neither agent can truly verify the intent of the other agent before they act. Thus, if a transaction takes place, it can be entirely attributed to trust since both agents needed to overcome the uncertainty associated with the transaction. Let us suppose, however, that the value of the good and the size of the payment are sufficiently high such that no amount of mutual trust allows a direct transaction to take place. To handle this situation, both agents could form a coalition with a mutually trusted third party, such as an escrow agent. The escrow agent would receive the payment from one agent to verify that the good can be shipped, and then later disperse the payment to the other agent (minus the escrow fee) when the good has been verified as received. Here, each agent benefits from the cooperative gains of the transaction. These gains would not be possible if even one agent chose to disband from the coalition.

This chapter intends to show how one could mathematically describe these types of trust-based interactions via the *cooperative trust game* to predict coalition formation and disbanding. It presents a rigorous treatment of coalition formation using cooperative game theory as the underlying mathematical framework. It is important to highlight that cooperative game theory is significantly different than the more widely recognized competitive (non-cooperative) game theory. Cooperative game theory focuses on what groups of self-interested agents can achieve. It is not concerned with how agents make choices or coordinate in coalitions, and does not assume that agents will always agree to follow arbitrary instructions. Rather, cooperative game theory defines games that tell how well a coalition can do for itself. And while the coalition is the basic modeling unit for coalition game, the theory supports modeling individual agent preferences without concern for their possible actions. As such, it is an ideal framework for modeling trust-based coalition formation since it can show how each agent's trust preferences can influence a group's ability to reason about trustworthiness. We refer the reader to [5] for an excellent primer on cooperative game theory.

2. Classes of trust games

This section characterizes different classes of trust games within the context of cooperative game theory. Our characterizations provide the necessary conditions for a coalition trust game to be classified into a particular class. We start with additive and constant-sum trust games,

which have limited value for cooperative applications, but are included for completeness. Then, we discuss superadditive and convex trust games, which show conditions for agents to form a grand coalition. In general, grand coalition solution concepts presented here can also be applied to smaller coalitions within a trust game through the use of a trust subgame.

2.1. Preliminaries

Let $\Gamma = (N, v)$ be a coalitional trust game with transferable utility where:

- N is a finite set of agents, indexed by i
- $v: 2^N \to \mathbb{R}$ associates with each coalition $S \subseteq N$ a real-valued payoff $v(S)$ that is distributed between the agents. Singleton coalitions, by definition, are assigned no value; i.e. $v(i) = 0 \; \forall i \in N$.

The transferable utility assumption means that payoffs in a coalition may be freely distributed among its members. With regards to payoff value of trust between agents, this assumption can be interpreted as a universal means for agents to mutually share the value of their trustworthy relationships. Trust cultivation often requires reciprocity between two agents as a necessary behavior to develop trust, and a transferable utility is a convenient way to model the exchange for this notion.

In defining a transferable payoff value of trust, one aspect to consider are the "goods of trust". These refer to opportunities for cooperative activity, knowledge, and autonomy. In this chapter, we refer to these goods as *trust synergy* $s(S)$, which is a trust-based result that could not be obtained independently by two or more agents. We may also interpret trust synergy as the value obtained by agents in a coalition as a result of being able to work together due to their attitudes of trust for each other. In defining a set function for trust synergy, it is important to explicitly show how each agent's attitude of trustworthiness for every other agent in a coalition affects this synergy. In general, higher levels of trust in a coalition should produce higher levels of synergy.

The payoff value of trust, however, also includes an opposing force in the form of vulnerability exposure, which we refer to as *trust liability* $l(S)$. Trusting involves being optimistic that the trustee will do something for the truster; and this optimism is what causes the vulnerability, since it restricts the inferences a truster makes about the likely actions of the trustee. However, the refusal to be vulnerable tends to undermine trust since it does not allow others to prove their own trustworthiness, stifling growth in trust synergy. Thus, we see that agents in trust-based relationships with other agents must be aware of the balance between the values of the trust synergy and trust liability in addition to their relative magnitudes.

Let the characteristic payoff function of a trust game be the difference between the trust synergy and trust liability of a coalition S.

$$v(S) = s(S) - l(S) \tag{1}$$

2.2. Additive trust game

Additive games are considered inessential games in cooperative game theory since the value of the union of two disjoint coalitions ($S_1 \cap S_2 = \emptyset$) is equivalent to the sum of the values of each coalition.

$$v(S_1 \cup S_2) = v(S_1) + v(S_2)\ \forall S_1, S_2 \subset N \tag{2}$$

We see that the total value of the trust relationships between any two disjoint coalitions must always be zero. In other words, the trust synergy between any two disjoint coalitions must always result in a value that is equal to their trust liability. Thus, by expanding this definition for trust games and rearranging the terms, we can characterize an additive trust game as:

$$\begin{gathered} s(S_1 \cup S_2) - l(S_1 \cup S_2) = s(S_1) - l(S_1) + s(S_2) - l(S_2) \\ \{\forall S_1, S_2 \subset N : S_1 \cap S_2 = \emptyset\} \end{gathered}$$

$$\begin{gathered} s(S_1 \cup S_2) - s(S_1) - s(S_2) = l(S_1 \cup S_2) - l(S_1) - l(S_2) \\ \{\forall S_1, S_2 \subset N : S_1 \cap S_2 = \emptyset\}) \end{gathered} \tag{3}$$

2.3. Constant-sum trust game

In constant-sum games, the sum of all coalition values in N remains the same, regardless of any outcome.

$$v(N) = v(S) + v(N \backslash S) = k\ \forall S \subset N \tag{4}$$

By expanding this definition for trust games and rearranging the terms, we can see that the constant-sum trust game is a special case of a two-coalition additive trust game involving every agent in the game.

$$\begin{gathered} s(N) - l(N) = s(S) - l(S) + s(N \backslash S) - l(N \backslash S)\ \forall S \subset N \\ s(N) - s(S) - s(N \backslash S) = l(N) - l(S) - l(N \backslash S)\ \forall S \subset N \end{gathered} \tag{5}$$

Definition: An agent is a *dummy agent* if the amount the agent contributes to any coalition is exactly the amount that it is able to achieve alone.

Theorem: Γ is a constant-sum trust game implies that Γ is a zero-sum trust game.

Proof: If Γ is a constant-sum game, the following constraint for singleton coalitions must always hold:

$$s(N) - s(i) - s(N \backslash i) = l(N) - l(i) - l(N \backslash i)\ \forall i \in N \tag{6}$$

By rearranging the terms, combining, and substituting, we get:

$$\begin{gathered} s(N) - l(N) = s(i) - l(i) + s(N \backslash i) - l(N \backslash i)\ \forall i \in N \\ v(N) = v(i) + v(N \backslash i)\ \forall i \in N \end{gathered}$$

$$v(N) = v(N\backslash i)\ \forall i \in N \tag{7}$$

The last equation implies that every agent in N must behave like a dummy agent if Γ is a constant-sum trust game. Since all agents behave like dummy agents and $v(i) = 0$ for all $i \in N$, then any coalition that forms in Γ will have no value. Hence, the value of the grand coalition is zero (i.e. $v(N) = k = 0$). Therefore, the only possible constant-sum trust game is the zero-sum trust game. This completes the proof.

Corollary: Γ is a zero-sum trust game if $s(S) = l(S)\ \forall S \subset N$.

Proof: If $s(S) = l(S)\ \forall S \subset N$, then $v(S) = 0\ \forall S \subset N$. Thus, $v(N) = v(N\backslash S) = k\ \forall S \subset N$. This result in implies that every possible coalition in N must behave like a coalition of dummy agents in a constant-sum trust game and their combinations with other coalitions will yield no value. Hence, the value of the grand coalition is always zero (i.e. $v(N) = k = 0$). This completes the proof.

Our proofs show that any constant-sum trust game is necessarily a zero-sum trust game that represents a special case of an additive trust game. These facts reinforce a notion that a group of agents who do not trust each other will always prefer to work as singleton coalitions. And even if there is some mutual trust between agents, gains from trust synergy are always lost to the trust liability, making it irrational to form any coalition with any other agent. Thus, if one determines that Γ is a constant-sum trust game, then this provides immediate justification for using non-cooperative game theory as the basis for modeling the purely competitive agents.

2.4. Superadditive trust game

In a superadditive game, the value of the union of two disjoint coalitions ($S_1 \cap S_2 = \emptyset$) is never less than the sum of the values of each coalition.

$$v(S_1 \cup S_2) \geq v(S_1) + v(S_2)\ \forall S_1, S_2 \subset N \tag{8}$$

This implies a monotonic increase in the value of any coalition as the coalition gets larger.

$$S \subseteq A \subseteq N \rightarrow v(S) \leq v(A) \leq v(N) \tag{9}$$

This property of superadditivity tells us that the new links that are established between the agents in the two disjoint coalitions are the sources of the monotonic increases. This results in a snowball effect that causes all agents in the game to form the grand coalition (a coalition containing all agents in the game) since the total value of the new trust relationships *between* any two disjoint coalitions must always be positive semi-definite. In other words, the trust synergy between any two disjoint coalitions must always result in a value that is at least as large as their combined individual trust liabilities. Thus, by expanding the definition for trust games and rearranging the terms, we can characterize a superadditive trust game as:

$$s(S_1 \cup S_2) - l(S_1 \cup S_2) \geq s(S_1) - l(S_1) + s(S_2) - l(S_2)$$
$$\{\forall S_1, S_2 \subset N : S_1 \cap S_2 = \emptyset\}$$

$$s(S_1 \cup S_2) - s(S_1) - s(S_2) \geq l(S_1 \cup S_2) - l(S_1) - l(S_2)$$
$$\{\forall S_1, S_2 \subset N : S_1 \cap S_2 = \emptyset\} \tag{10}$$

2.5. Convex trust games

A game is convex if it is supermodular, and this trivially implies superadditivity (when $S_1 \cap S_2 = \emptyset$). Thus, we see that convexity is a stronger condition than superadditivity since the restriction that two coalitions must be disjoint no longer applies.

$$v(S_1 \cup S_2) + v(S_1 \cap S_2) \geq v(S_1) + v(S_2)\ \forall S_1, S_2 \subset N \tag{11}$$

In convex games, the incentives of joining a coalition grow as the coalition gets larger. This means that the marginal contribution of each agent $i \in N$ is non-decreasing.

$$v(S \cup i) - v(S) \leq v(A \cup i) - v(A) \text{ whenever } S \subset A \subset N \backslash i \tag{12}$$

Definition: A *subgame* $v_R : 2^R \to \mathbb{R}$, where $R \subseteq N$ is not empty, is defined as $v_R(S) = v(S)$ for each $S \subseteq N$. In general, solution concepts that apply to a grand coalition can also apply to smaller coalitions in terms of a subgame.

Definition: Given a game $\Gamma = (N, v)$ and a coalition $R \subseteq N$, the R-marginal game $v_R : 2^{N \backslash R} \to \mathbb{R}$ is defined by $v_R(S) = v(R \cup S) - v(R)$ for each $S \subseteq N \backslash R$.

Using these definitions, Branzei, Dimitrov, and Tijs proved that a game is convex if and only all of its marginal games are superadditive [6]. We provide their proof here as a means for the reader to readily justify this assertion.

Theorem:

A game $\Gamma = (N, v)$ is convex if and only if for each $R \in 2^N$ the *R-marginal game* $(N \backslash R, v_R)$ is superadditive.

Proof: Suppose (N, v) is convex. Let $R \subseteq N$ and $S_1, S_2 \subseteq N \backslash R$. Then:

$$\begin{aligned} v_R(S_1 \cup S_2) + v_R(S_1 \cap S_2) &= v(R \cup S_1 \cup S_2) + v(R \cup (S_1 \cap S_2)) - 2v(R) \\ &= v\big((R \cup S_1) \cup (R \cup S_2)\big) + v\big((R \cup S_1) \cap (R \cup S_2)\big) - 2v(R) \\ &\geq v(R \cup S_1) + v(R \cup S_2) - 2v(R) \\ &= \big(v(R \cup S_1) - v(R)\big) + \big(v(R \cup S_2) - v(R)\big) \\ &= v_R(S_1) + v_R(S_2) \end{aligned} \tag{13}$$

where the inequality follows from the convexity of v. Hence, v_R is convex (and superadditive as well).

Now, let $S_1, S_2 \subseteq N$ and $R = S_1 \cap S_2$. Suppose that for each $R \in 2^N$, the game $(N \backslash R, v_R)$ is superadditive. If $R = \emptyset$, then the game $(N \backslash \emptyset, v_\emptyset) = (N, v)$ and $v(\emptyset) = 0$; hence, Γ is superadditive. If $R \neq \emptyset$, then because $(N \backslash R, v_R)$ is superadditive:

$$v_R((S_1 \cup S_2) \backslash R) \geq v_R(S_1 \backslash R) + v_R(S_2 \backslash R)$$
$$v(S_1 \cup S_2) - v(R) \geq\ v(S_1) - v(R) + v(S_2) - v(R)$$

$$\begin{aligned} v(S_1 \cup S_2) + v(R) &\geq v(S_1) + v(S_2) \\ v(S_1 \cup S_2) + v(S_1 \cap S_2) &\geq v(S_1) + v(S_2) \end{aligned} \tag{14}$$

This completes the proof.

By using this characterization in the previous theorem and expanding it to our definition of a trust game, we can state a necessary requirement to produce a convex trust game: that the marginal trust synergy between any two coalitions must always result in a value that is at least as large as their marginal trust liability.

$$s_R((S_1 \cup S_2)\backslash R) - l_R((S_1 \cup S_2)\backslash R) \geq s_R(S_1\backslash R) - l_R(S_1\backslash R) + s_R(S_2\backslash R) - l_R(S_2\backslash R)$$
$$\{\forall S_1, S_2 \subset N : S_1 \cap S_2 = R\}$$

$$s_R((S_1 \cup S_2)\backslash R) - s_R(S_1\backslash R) - s_R(S_2\backslash R) \geq l_R((S_1 \cup S_2)\backslash R) - l_R(S_1\backslash R) - l_R(S_2\backslash R) \tag{15}$$
$$\{\forall S_1, S_2 \subset N : S_1 \cap S_2 = R\}$$

3. Trust game model

In the previous section, we characterized different classes of trust games without explicitly defining a trust game model. In this section, we provide a general model for trust games that conforms to the theoretical constructions in the previous section and can be adapted to a wide variety of applications.

3.1. Modeling trust synergy and trust liability

The attitude of trustworthiness agents have toward other agents in a trust game is managed in an $|N| \times |N|$ matrix T.

$$T = \left[t_{i,j}\right]_{|N|\times|N|} = \begin{cases} t_{i,j} = 1, & i = j \\ t_{i,j} \in [0,1], & i \neq j \end{cases} \tag{16}$$

This matrix is populated with values $t_{i,j}$ that represent the probability that agent j is trustworthy from the perspective of agent i. The values $t_{i,j}$ can also be interpreted as the probabilities that agent i will allow agent j to interact with him, since rational agents would prefer to interact with more trustworthy agents.

The manner in which $t_{i,j}$ is evaluated depends on an underlying trust model. We make no assumption about the use of a particular trust model, as the choice of an appropriate model may be application-specific. We also make no assumption about the spatial distribution of the agents in a game – therefore, this matrix should not necessarily imply the structure of a communications graph.

We provide a general model for trust synergy and trust liability that can be adapted for a variety of applications. Our model makes use of a symmetric matrix Σ to manage potential trust synergy and a matrix Λ to manage potential trust liability. Σ is symmetric because we assume that agents mutually agree on the benefits of a synergetic interaction.

$$\Sigma = [\sigma_{i,j}]_{|N|\times|N|} = \begin{cases} \sigma_{i,j} = 0, & i = j \\ \sigma_{i,j} = \sigma_{j,i} \geq 0, & i \neq j \end{cases} \tag{17}$$

$$\Lambda = [\lambda_{i,j}]_{|N|\times|N|} = \begin{cases} \lambda_{i,j} = 0, & i = j \\ \lambda_{i,j} \geq 0, & i \neq j \end{cases} \tag{18}$$

As with the T matrix, we make no assumptions about how Σ and Λ are calculated, since the meaning of their values may depend on the application. For example, the calculations for $\sigma_{i,j}$ and $\lambda_{i,j}$ between two agents may not only take into account each agent's individual intrinsic attributes – it may also factor in externalities (i.e. political climate, weather conditions, pre-existing conditions, etc.) that neither agent has direct control over.

Definition: The total value of the trust synergy in a coalition is defined as the following set function:

$$s(S) = \sum_{i,j\in S} \sigma_{i,j}\, t_{i,j} t_{j,i} \;\forall i > j \tag{19}$$

Trust synergy is the value obtained by agents in a coalition as a result of being able to work together due to their attitudes of trust for each other. The set function $s(S)$ assumes that the events "agent i allows agent j to interact" and "agent j allows agent i to interact" are independent. This is reasonable since agents are assumed to behave as independent entities within a trust game (i.e. no agent is controlled by any other agent). Therefore, we treat the product $t_{i,j}t_{j,i}$ as the relative strength of a trust-based synergetic interaction, which justifies the use of the summation. The value for $\sigma_{i,j}$ serves as a weight for a trust-based synergetic interaction.

Definition: The total value of the trust liability in a coalition is defined as the following set function:

$$l(S) = \sum_{i,j\in S} \lambda_{i,j} t_{i,j} \;\; \forall i \neq j \tag{20}$$

Trust liability can be thought of as the vulnerability that agents in a coalition expose themselves to due to their attitudes of trust for each other. We treat the product $\lambda_{i,j}t_{i,j}$ as a measure for agent i's exposure to unfavorable trust-based interactions from agent j. A high amount of trust can expose agents to high levels of vulnerability. But each agent can regulate its exposure to trust liability by adjusting $t_{i,j}$. Changes to $t_{i,j}$, however, also influence the benefits of trust synergy.

3.2. Modeling the trust game

We define the trust game (also known as the total value of the trust payoff in a coalition) as the difference between its trust synergy and trust liability.

$$\begin{aligned} v(S) &= \sum_{\substack{i,j\in S \\ \forall i>j}} \sigma_{i,j}\, t_{i,j} t_{j,i} - \sum_{\substack{i,j\in S \\ \forall i\neq j}} \lambda_{i,j} t_{i,j} \\ v(S) &= \sum_{\substack{i,j\in S \\ \forall i>j}} t_{i,j} t_{j,i} \left(\sigma_{i,j} - \frac{\lambda_{i,j}}{t_{j,i}} - \frac{\lambda_{j,i}}{t_{i,j}}\right) \end{aligned} \tag{21}$$

The factorization shows us that the first factor ($t_{i,j}t_{j,i}$) will always be greater than or equal to zero while the second factor can be either positive or negative. Hence, by isolating the second factor and recognizing that trust values equal to 1 produce the smallest possible reduction in the second factor, we can state the condition that guarantees the potential for two agents to form a trust-based pair coalition.

Proposition 1: Any two agents $i, j \in N$ will never form a trust-based pair coalition if $\sigma_{i,j} < \lambda_{i,j} + \lambda_{j,i}$. Otherwise, the potential exists for agent i and j to form a trust-based pair coalition.

Proposition 2: If two agents can never form a trust-based pair coalition, then the best strategy for both agents is to never trust each other (i.e. $t_{i,j} = t_{j,i} = 0$).

In general, proposition 1 does not extend to trust-based coalitions larger than two due to the complex coupling of trust dynamics between different agents as coalitions grow larger. For example, two agents who may produce a negative trust payoff value as a pair may actually realize a positive trust payoff with the addition of a third agent. This situation occurs if both agents have positive trust relationships with the third agent that outweighs their own negative trust relationship. Such a situation is common in real world scenarios, and justifies the importance of various trusted third parties, such as escrow companies, website authentication services, and couples therapists.

In light of this, we can mathematically justify a condition similar to proposition 1 that is valid for coalitions of any size – but only for a special type of trust game.

Theorem: A trust-based coalition $S \subseteq N$ will never form if:

$$\sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} < \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}} \tag{22}$$

$$\{\forall i, j \in S : t_{i,j}t_{j,i} = k\}$$

Proof: Let $S \subseteq N$ and $t_{i,j}t_{j,i} = k$ for all $i, j \in S$. Then, by substituting k into the trust model:

$$v(S) = \sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j}\, k - \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}} k \tag{23}$$

$$v(S) = k\left(\sum_{\substack{i,j \in S \\ \forall i > j}} \sigma_{i,j} - \sum_{\substack{i,j \in S \\ \forall i \neq j}} \frac{\lambda_{i,j}}{t_{j,i}}\right)$$

Because k is a constant that is always greater than or equal to zero, we can clearly see that the second factor affects whether or not $v(S)$ is positive or negative. Hence, if the second term in the second factor is larger than the first term in the second factor, then a coalition S will never form. This completes the proof.

3.3. Incorporating context into a trust game

In practice, trust is often defined relative to some context. Context allows individuals to simplify complex decision-making scenarios by focusing on more narrow perspectives of situations or others, avoiding the potential for inconvenient paradoxes.

Coalitional trust games can also be defined relative to different contexts using the multi-issue representation [7], where we use the words "context" and "issue" interchangeably.

Definition: A multi-issue representation is composed of a collection of coalitional games, each known as an issue, $(N_1, v_1), (N_2, v_2), \cdots, (N_k, v_k)$, which together constitute the coalitional game (N, v) where

- $N = N_1 \cup N_2 \cup \cdots \cup N_k$
- For each coalition $S \subseteq N$, $v(S) = \sum_{i=1}^{k} v_i(S \cap N_i)$

This approach allows us to define an arbitrarily complex trust game that can be easily decomposed into simpler trust games relative to a particular context. A set of agents in one context can overlap partially or complete with another set of agents in another context. And one can choose to treat the coalitional game in one big context, or the union of any number of contexts based on some decision criteria.

3.4. Altruistic and competitive contribution decomposition

In the analysis of a trust-based coalition, it may sometimes be useful to understand the manner in which different subsets of a coalition contribute to its payoff value. One way to do this is to use a framework developed by Arney and Peterson where measures of cooperation are defined in terms of altruistic and competitive cooperation [8]. The unifying concept in the framework is a *subset team game*, a situation or scenario in which the value of a given outcome (as perceived by a team subset) can be measured.

Definition: Given a game $\Gamma = (N, u)$ and a non-empty coalition $R \subseteq S \subseteq N$, the *subset team game* $u_R: 2^R \to \mathbb{R}$ associates a valued payoff $u_R(S)$ perceived by the agents in R when the agents in S cooperate.

The authors limit the application of the framework to games where more agents in a coalition lead to more successful outcomes. Thus, adding more agents to a coalition should never reduce the coalition's payoff value. Also, the payoff value perceived by a coalition should not be smaller than the payoff value perceived by a subset of the same coalition. We refer to these two properties as *fully-cooperative* and *cohesive*, respectively.

Definition: A subset team game is fully-cooperative if $u_A(B) \leq u_A(C)$ for all $A \subseteq B \subseteq C \subseteq N$.

Definition: A subset team game is *cohesive* if $u_A(C) \leq u_B(C)$ for all $A \subseteq B \subseteq C \subseteq N$.

The authors show that in a fully-cooperative and cohesive game, the marginal contribution of a subset team is equal to the sum of the competitive and altruistic contributions of the subset team.

Definition: Given a payoff function $u_R(S)$ in a subset team game, the *total marginal contribution* of $R \subseteq S$ to a team S is $m_R(S) = u_S(S) - u_{S\backslash R}(S\backslash R)$. If the game is both cohesive and fully-cooperative, then the *competitive contribution* of R is $c_R(S) = u_S(S) - u_{S\backslash R}(S)$ and

the *altruistic contribution* is $a_R(S) = u_{S\backslash R}(S) - u_{S\backslash R}(S\backslash R)$. Note that the total marginal contribution decomposes as $m_R(S) = c_R(S) + a_R(S)$.

In order to use these definitions within a trust game, we must first show they relate to the coalition game classes described in Section 3.

Theorem: A subset team game that is both fully-cooperative and cohesive is a convex game.

Proof:

First, we prove the fully-cooperative case. If $u_A(B) \leq u_A(C)$ such that $A \subseteq B \subseteq C \subseteq N$, then following inequalities are also true:

$$u_A(B) \leq u_A(B \cup i) \qquad A \subseteq B \subseteq N\backslash i \tag{24}$$

$$u_A(C) \leq u_A(C \cup i) \qquad A \subseteq C \subseteq N\backslash i \tag{25}$$

$$u_A(B \cup i) \leq u_A(C \cup i) \qquad A \subseteq B \subseteq C \subseteq N\backslash i \tag{26}$$

Since the system of inequalities shows that the contribution of an additional agent in a coalition is always non-decreasing, it is trivially true that:

$$u_A(B \cup i) - u_A(B) \leq u_A(C \cup i) - u_A(C)\ A \subseteq B \subseteq C \subseteq N\backslash i \tag{27}$$

Next, we prove the cohesive case. If $u_A(C) \leq u_B(C)$ such that $A \subseteq B \subseteq C \subseteq N$, then following inequalities are also true:

$$u_A(C) \leq u_{A\cup i}(C) \qquad A \subseteq C \subseteq N\backslash i \tag{28}$$

$$u_B(C) \leq u_{B\cup i}(C) \qquad B \subseteq C \subseteq N\backslash i \tag{29}$$

$$u_{A\cup i}(C) \leq u_{B\cup i}(C) \qquad A \subseteq B \subseteq C \subseteq N\backslash i` \tag{30}$$

Since the system of inequalities show that the contribution of an additional agent in the accessing coalition subset is always non-decreasing, it is trivially true that:

$$u_{A\cup i}(C) - u_A(C) \leq u_{B\cup i}(C) - u_B(C)\ A \subseteq B \subseteq C \subseteq N\backslash i \tag{31}$$

This completes the proof.

It is important to note that the additional agent i for both cases is never already inside either coalition B or C. If it was, then the proof would be invalid, as one could easily demonstrate counter examples under cases where an agent $i \in C\backslash B$.

Now that we have shown that a convex subset team game is fully-cooperative and cohesive, we may decompose the total marginal contribution of a set of agents into both altruistic and competitive contributions whenever a trust game is convex. To do so, we must define a value function $u_R(S)$ that utilizes the trust game payoff value function $v(S)$.

Definition: Given a game $\Gamma = (N, u)$ and a non-empty coalition $R \subseteq S \subseteq N$, the *subset trust game* $u_R: 2^R \to \mathbb{R}$ associates a trust payoff value $u_R(S)$ perceived by the agents in R when the agents in S cooperate:

$$u_R(S) = v(R) + \sum_{i \in R, j \in S \setminus R} v(\{i, j\}) \; R \subseteq S \subseteq N \tag{32}$$

The rationale behind this payoff function in is that the payoff has to be from the perspective of the agents in R. The agents in R can factor in the values related to relationships between themselves (first term) and relationships between agents in R and agents in S (second term). But they cannot factor in values related to relationships between the agents in $S \setminus R$, since agents in R are assumed to have no direct knowledge of what is happening between the $S \setminus R$ agents.

Using the payoff function $u_R(S)$, we can calculate R's altruistic contribution $a_R(S)$ and competitive contribution $c_R(S)$ in a coalition S.

$$a_R(S) = \sum_{i \in R, j \in S \setminus R} v(\{i, j\}) \; R \subseteq S \subseteq N \tag{33}$$

$$c_R(S) = v(S) - v(S \setminus R) - \sum_{i \in R, j \in S \setminus R} v(\{i, j\}) \; R \subseteq S \subseteq N \tag{34}$$

$$m_R(S) = a_R(S) + c_R(S) = v(S) - v(S \setminus R) \; R \subseteq S \subseteq N \tag{35}$$

4. Convoy trust game

In this section, we present an example of cooperative trust for a specific application: the convoy. Our primary purpose here is to demonstrate how one could use the theory in this chapter to model specific scenarios involving trust. We define the *convoy trust game*, which describes a cooperative game where the agents intend to move forward together in a single file. This type of game can be naturally adapted to the analysis of traffic patterns, leader-follower applications, hierarchical organizations, or applications with sequential dependencies. Our goal in this section is to understand how trust-based coalitions will form under this type of scenario.

4.1. 4-Agent convoy case

We begin with a simple convoy scenario that models a four-agent convoy, $N = \{1,2,3,4\}$, which intends to move together in a single file. The value of each index into N represents the agent's position in the convoy. For this scenario, we interpret the trust synergy in the coalition to represent the agents in the coalition moving forward. Thus, we set the values in the trust synergy matrix Σ equal to the number of agents that will move forward if the two agents are moving forward (inclusive of the two agents). We interpret the trust liability in coalition to represent the vulnerability of agents in the coalition to stop moving. Thus, we set the values in the trust liability matrix Λ equal to the number of agents that can prevent a particular agent from moving forward in a agent coalition pair.

Definition: The values in Σ and Λ for a 4-convoy trust game are:

$$\Sigma = \begin{bmatrix} 0 & 2 & 3 & 4 \\ 2 & 0 & 3 & 4 \\ 3 & 3 & 0 & 4 \\ 4 & 4 & 4 & 0 \end{bmatrix} \Lambda = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 2 & 2 & 0 & 2 \\ 3 & 3 & 3 & 0 \end{bmatrix}$$

4.2. 4-Agent convoy trust game analysis

First, let us analyze this game as an additive trust game. While there are infinitely many solutions for T that conform to the additive game, the most obvious solution is the extreme situation where no agent trusts any other agent – or, when T is the identity matrix ($T = I$). In this case, no agent will ever affect another agent, either positively or negatively. Thus, each agent will ultimately form a singleton coalition and fail to work cooperatively with any other agent.

Next, let us analyze another extreme situation where every agent completely trusts every other agent – or, when $T = [1]_{4\times4}$. As such, we can enumerate the trust payoff values for each possible coalition.

$$v(\{1,2\}) = 1;\ v(\{1,3\}) = 1;\ v(\{1,4\}) = 1;\ v(\{2,3\}) = 0;\ v(\{2,4\}) = 0;$$

$$v(\{3,4\}) = -1;\ v(\{1,2,3\}) = 2;\ v(\{1,2,4\}) = 2;\ v(\{1,3,4\}) = 1;$$

$$v(\{2,3,4\}) = -1;\ v(\{1,2,3,4\}) = 2;$$

These results provide us an interesting insight, in that all agents behind the lead agent find higher values of trust payoff with the lead agent than with the nearest agent. As such, as long as the lead agent is a member of a trust-based coalition in this game, there will be no incentive for any other agent to abandon the coalition. Thus, the agents ultimately form the grand coalition. Note, however, that the formation of a grand coalition does not imply that the trust game is superadditive or convex. This assertion is justified with the observation that $v(\{3,4\}) \not\geq v(\{3\}) + v(\{4\}) = 0$.

In order to form a convex 4-convoy trust game, we must satisfy the conditions that ensure that all trust payoff values in any coalition are at least as large as any sub-coalition – or that the marginal trust synergy is always greater than or equal to the marginal trust liability. While, again, there are infinitely many solutions for T that conform to convex game, the games with the highest trust payoff actually have either one of the following trust matrices (see next section for proof):

$$T_1 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} T_2 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix} T_3 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{bmatrix} T_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \end{bmatrix}$$

T_1, T_2, T_3, and T_4 are modified versions of $[1]_{4\times4}$ and all produce the same results in the trust payoff value function. The main modification ensures that agents 3 and 4 have no trust

toward each other since the sum of their individual trust liabilities always outweigh the trust synergy they create. The following is the enumeration of the trust payoff values for the 4-convoy trust game with the highest trust payoff:

$$v(\{1,2\}) = 1;\ v(\{1,3\}) = 1; v(\{1,4\}) = 1;\ v(\{2,3\}) = 0;\ v(\{2,4\}) = 0;$$

$$v(\{3,4\}) = 0; v(\{1,2,3\}) = 2;\ v(\{1,2,4\}) = 2;\ v(\{1,3,4\}) = 2;$$

$$v(\{2,3,4\}) = 0;\ v(\{1,2,3,4\}) = 3;$$

The deep insight we gain from analyzing optimal trust matrices and payoff value results is that all agents behind the lead agent need only trust the lead agent in the convoy to move forward, provided the lead agent trusts every other agent to follow it. This echoes the intuition seen in Jean-Jacques Rousseau's classic "stag hunt" game, where there is no incentive for any player to cheat by not cooperating as long as each player can trust others to do the same [9].

We can use anecdotal evidence found in our experiences in automobile traffic jams to verify our understanding of the theoretical result. Drivers in traffic lanes (coalitional convoys) rarely place a significant amount of trust in neighboring drivers to justify the value of the traffic lane (as the model corroborates). In fact, in the event a driver becomes stuck in a traffic jam, he likely will not feel betrayed by the driver directly in front. Instead, he will unconsciously begin gauging the coalitional value of the traffic jam by considering his level of trust in the lead driver in the traffic jam, whether in visible range or not. In most cases, the driver monitors the traffic flow or listens to traffic reports to gauge his trust for the lead driver. He may also unconsciously consider other drivers in the traffic jam and estimate their trust perceptions of the traffic jam to gauge the coalition's value. In the event a driver cannot accurately gauge the value of the traffic jam, he may choose to leave the traffic jam and attempt to join another traffic coalition (lane) with a higher payoff value. These types of driver behaviors are generally not performed when the trust for the lead driver to move forward is high. Yet, these behaviors feel necessary when the trust lessens since they attempt to resolve coalitional and environmental uncertainties.

4.3. N-Agent convoy optimal solution proof

We conclude by generalizing the convoy trust game for any number of agents and prove the solution for the highest payoff trust-based coalition. Our proof shows that all agents behind the lead agent in a convoy need only trust the lead agent, and no other agent, to move forward so long as the lead agent trusts every other agent to follow it.

Definition: The generalized values in Σ and Λ for a convoy trust game with $|N|$ agents are:

$$\Sigma = \left[\sigma_{i,j}\right]_{|N|\times|N|} = \begin{cases} \sigma_{i,j} = 0, & i = j \\ \sigma_{i,j} = \max(\{i,j\}), & i \neq j \end{cases} \tag{36}$$

$$\Lambda = \left[\lambda_{i,j}\right]_{|N|\times|N|} = \begin{cases} \lambda_{i,j} = 0, & i = j \\ \lambda_{i,j} = i - 1, & i \neq j \end{cases} \tag{37}$$

Theorem: The convoy trust game that produces the grand coalition with highest payoff value has a trust matrix that conforms to the following construction:

$$T = \left[t_{i,j}\right]_{|N|\times|N|} = \begin{cases} t_{i,j} = 1, & i = j \\ t_{i,j} = 1, & i \neq j, \min(\{i,j\}) = 1 \\ t_{i,j} = t_{j,i} \in \{0,1\}, & i \neq j, \min(\{i,j\}) = 2 \\ t_{i,j} = 0, & i \neq j, \min(\{i,j\}) > 2 \end{cases} \tag{38}$$

Proof:

Suppose we generalize the values in Σ and Λ. According to proposition 1, two agents $i, j \in N$ will never form a trust-based coalition pair if $\sigma_{i,j} < \lambda_{i,j} + \lambda_{j,i}$. Thus, by substitution:

$$\max(\{i,j\}) < (i-1) + (j-1)$$

$$\max(\{i,j\}) < i + j - 2 \tag{39}$$

We see that if i is the maximum value, then $0 < j - 2$. Similarly, if j is the maximum value, then $0 < i - 2$. Thus, the inequalities tell us that any agent behind the second agent will never form a trust-based coalition with any other agent behind the second agent. Therefore, by proposition 2, the best strategy for these agents is to have no trust for each other; hence $t_{i,j} = 0$ when $\min(\{i,j\}) > 2$ for $i \neq j$.

The equalities above alsto tell us that a trust-based coalition formation is possible with the lead agent and the second agent. Using the definition of the trust game model, the trust payoff values for a coalition in the convoy trust game is:

$$v(S) = \sum_{\substack{i,j\in S \\ \forall i>j}} t_{i,j}t_{j,i}\left(\max(\{i,j\}) - \frac{i-1}{t_{j,i}} - \frac{j-1}{t_{i,j}}\right) \tag{40}$$

We may now define the trust payoff values for any pair of agents as:

$$v(\{i,j\}) = t_{i,j}t_{j,i}\left(\max(\{i,j\}) - \frac{i-1}{t_{j,i}} - \frac{j-1}{t_{i,j}}\right) \tag{41}$$

Let us first analyze coalition formation with the lead agent. If $i = 1$, then $\max(\{i,j\}) = j$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{1,j\}) = t_{1,j}t_{j,1}\left(j - \frac{j-1}{t_{1,j}}\right)$$

$$v(\{1,j\}) = jt_{1,j}t_{j,1} - jt_{j,1} + t_{j,1}$$

$$v(\{1,j\}) = t_{j,1}\left(jt_{1,j} - j + 1\right) \tag{42}$$

By inspection, we see that the highest trust payoff value is achieved when both the lead agent and any other agent completely trust each other (i.e., when $t_{1,j} = t_{j,1} = 1$). However, to justify this assertion, we must also show this is true when $j = 1$. If $j = 1$, then $\max(\{i, j\}) = i$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{i,1\}) = t_{i,1}t_{1,i}\left(i - \frac{i-1}{t_{1,i}}\right)$$

$$v(\{i,1\}) = it_{i,1}t_{1,i} - it_{i,1} + t_{i,1}$$

$$v(\{i,1\}) = t_{i,1}\left(it_{1,i} - i + 1\right) \tag{43}$$

Again, by inspection, we confirm that the highest trust payoff is achieved when both the lead agent and any other agent completely trust each other. Therefore, $t_{i,j} = 1$ when the $\min(\{i, j\}) = 1$ for $i \neq j$.

Now, we analyze coalition formation with the second agent. If $i = 2$, then $\max(\{i, j\}) = j$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{2,j\}) = t_{2,j}t_{j,2}\left(j - \frac{1}{t_{j,2}} - \frac{j-1}{t_{2,j}}\right)$$

$$v(\{2,j\}) = t_{2,j}t_{j,2}j - t_{2,j} - jt_{j,2} + t_{j,2}$$

$$v(\{2,j\}) = t_{j,2}\left(jt_{2,j} - j + 1\right) - t_{2,j} \tag{44}$$

The highest trust payoff that can be achieved with the second agent is equal to zero, and this only occurs when both agents either have complete trust in each other (i.e., when $t_{2,j} = t_{j,2} = 1$) or no trust in each other (i.e., when $t_{2,j} = t_{j,2} = 0$). Any other combination of trust values will produce negative trust payoff values. However, to justify this assertion, we must also show this is true when $j = 2$. If $j = 2$, then $\max(\{i, j\}) = i$. Therefore, the payoff value for a pair coalition between i and j is:

$$v(\{i,2\}) = t_{i,2}t_{2,i}\left(i - \frac{i-1}{t_{2,i}} - \frac{1}{t_{i,2}}\right)$$

$$v(\{i,2\}) = it_{i,2}t_{2,i} - it_{i,2} + t_{i,2} - t_{2,i}$$

$$v(\{i,2\}) = t_{i,2}\left(it_{2,i} - i + 1\right) - t_{2,i} \tag{45}$$

By inspection, we confirm that the highest trust payoff that can be achieved with the second agent is equal to zero. Therefore, $t_{i,j} = t_{j,i} \in \{0,1\}$ when $\min(\{i, j\}) = 2$ for $i \neq j$.

To complete the proof, we simply state our assumption that each agent fully trusts itself, since it is impossible for an agent to diverge from a singleton coalition. Therefore, $t_{i,j} = 1$ when $i = j$. This completes the proof.

Author details

Dariusz G. Mikulski
U.S. Army Tank-Automotive Research Development and Engineering Center (TARDEC),
Warren, MI,
USA

Acknowledgement

The author would like to acknowledge the Ground Vehicle Robotics group at the U.S. Army Tank-Automotive Research Development, and Engineering Center (TARDEC) in Warren, MI for their basic research investment, which resulted in the development of the cooperative trust game theory in this chapter. Furthermore, the author would like to thank his academic advisors, Dr. Edward Gu (Oakland University) and Dr. Frank Lewis (University of Texas in Arlington) for their insight and advice, which tremendously helped to guide the research to a successful outcome.

5. References

[1] N. Luhmann, "Familiarity, Confidence, Trust: Problems and Alternatives," *Trust: Making and Breaking Cooperative Relations*, pp. 94-108, 1988.
[2] B. Adams and R. Webb, "Trust in Small Military Teams," in *7th International Command and Control Technology Symposium*, 2002.
[3] C. McLeod. (2006) Trust. [Online]. http://plato.stanford.edu/entries/trust/
[4] S. D. Ramchurn, D. Huynh, and N. R. Jennings, "Trust in Multi-Agent Systems," *The Knowledge Engineering Review*, vol. 19, no. 1, pp. 1-25, 2004.
[5] Y. Shoham and K. Leyton-Brown, "Teams of Selfish Agents: An Introduction to Coalitional Game Theory," in *Multiagent Systems: Algorithmic, Game-Theoretic, and Logical Foundations*. Cambridge: Cambridge University Press, 2009, pp. 367-391.
[6] R. Branzei, D. Dimitrov, and S. Tijs, "A new characterization of convex games," in *Tiburg University, Center of Economic Research*, 2004.
[7] V. Conitzer and T. Sandholm, "Computing Shapley values, manipulating value division schemes, and checking core membership in multi-issue domains," in *AAAI Conference on Artificial Intelligence*, 2004.
[8] D. C. Arney and E. Peterson, "Cooperation in social networks: communication, trust, and selflessness.," in *26th Army Science Conference*, Orlando, FL, 2008.

[9] A. Dixit and B. Nalebuff, "Prisoners' Dilemmas and How to Resolve Them," in *The Art of Strategy*. New York: W.W. Norton and Company, 2008, pp. 64-101.