# Strategic Coalitions in Systems with Catastrophic Failures

**Pavel Naumov**
Department of Mathematical Sciences
Claremont McKenna College
Claremont, California 91711
pnaumov@cmc.edu

**Kevin Ros**
Department of Computer Science
Vassar College
Poughkeepsie, New York 12604
kevinros@vassar.edu

## Introduction

In this paper we study failures of strategic coalitions. There are at least two forms of failures associated with strategies that we illustrate using an example of an airplane landing. On one hand, consider an airplane pilot and co-pilot coordinating their efforts for the flight to arrive on time. They might have a strategy to land the plane on time with certain probability. If they follow this strategy, but the plane arrives late, then we call this *failure of the strategy*. On the other hand, consider the same two agents (pilot and co-pilot) planning to land the airplane at one of the two airports. If both agents coordinate their actions, the plane is guaranteed to land at the airport of their choice, unless there is a *catastrophic system failure* such as a mechanical malfunction of the plane or a terrorist attack.
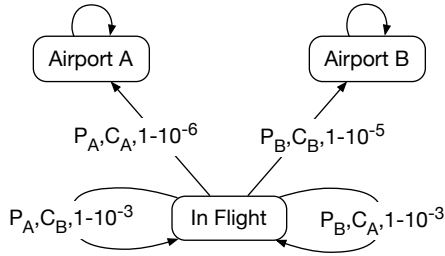


Figure 1: A Transition System.

The focus of this paper is on strategies of coalitions that are guaranteed to succeed unless catastrophic system failure occurs. We capture the setting of such strategies by transition systems with catastrophic failures. Figure 1 depicts an example of such a transition system for cooperation between a pilot and a co-pilot. Here, both the pilot and the co-pilot have strategies to land the plane at airport $A$. We call such strategies $P_A$ and $C_A$ respectively. They also have strategies $P_B$ and $C_B$ to land the plane at airport $B$. If the pilot chooses strategy $P_A$ and the co-pilot chooses strategy $C_A$, then the plane will land at the airport $A$ with probability $1 - 10^{-6}$ where one millionth accounts for the chance of a catastrophic failure. In this example, we assume that if the pilot

pilot and the co-pilot choose mismatching strategies, then the plane will remain in flight with a $10^{-3}$ chance of a catastrophic failure. We say that the coalition $\{P, C\}$, consisting of the pilot $P$ and the co-pilot $C$, has a strategy to land the plane at airport $A$ with *probability of survival* $1 - 10^{-6}$. We formally write this as $[\{P, C\}]_{1-10^{-6}}$("land at airport $A$").

Perhaps a more interesting example of a transition system is depicted in Figure 2. Here the pilot has two strategies for each airport: a regular and a quick. The latter is denoted by letter $Q$ instead of $P$. A quicker strategy allows the pilot to land the plane faster at a price of a higher chance of a catastrophic failure. In this example, the coalition $\{P, C\}$ still has a strategy to land the plane at airport $A$ with probability of survival $1 - 10^{-6}$. Additionally, the same coalition has a strategy to land the plane at airport $A$ quickly with probability of survival $1 - 10^{-5}$.
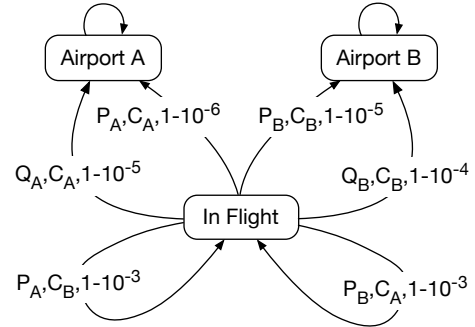


Figure 2: A Transition System.

In this paper we propose a logical system that describes universal properties of modality $[C]_p$ in all transition systems with catastrophic failures. The main technical contribution of this paper are the soundness and the completeness theorems for the proposed logical system. Axioms for the coalition power modality $[C]$ without probability of success were first proposed by (Pauly 2001; 2002).

## Syntax and Semantics

We assume a fixed finite set of agents $A$ and a fixed set of propositional variables. Additionally, a *coalition* is any nonempty subset of $A$.

**Definition 1** *Let $\Phi$ be the minimal set of formulae such that*

1. *$p \in \Phi$ for each propositional variable $p$,*
2. *$\neg\varphi, \varphi \to \psi \in \Phi$ for all formulae $\varphi, \psi \in \Phi$,*
3. *$[C]_p\varphi \in \Phi$ for each coalition $C$, each real number $p$ such that $0 \le p \le 1$, and each formula $\varphi \in \Phi$.*

We assume that Boolean constant $\top$ is defined in our language in the standard way.

Let $X^Y$ be the set of all functions from set $Y$ to set $X$. Below we define the notion of a transition system with catastrophic failures. Informally, at each state $s \in S$ of this system each agent $a \in A$ chooses an action from a domain of actions $D$. Function $P(s, \delta)$ specifies probability of survival of the system in state $s$ under complete strategy profile $\delta$. Assuming the system survives, action aggregation mechanism $M$ specifies possible next states based on state $s$ and a *complete action profile* $\delta \in D^A$.

**Definition 2** *A transition system with catastrophic failures is a tuple $(S, D, P, M, \pi)$, where*

1. *$S$ is a set (of states),*
2. *$D$ is a nonempty set (domain of actions),*
3. *$P$ is a function from set $S \times D^A$ into set $[0, 1]$,*
4. *$M \subseteq S \times D^A \times S$, where, for each state $s \in S$ and each complete action profile $\delta \in D^A$, if $P(s, \delta) > 0$, then there is at least one state $s' \in S$ such that $(s, \delta, s') \in M$.*
5. *$\pi$ is a mapping of propositional variables to subsets of $S$.*

Setting aside the probabilistic component, our transition system is slightly more general than (Pauly 2001; 2002) because we allow transitions to be non-deterministic. That is, for any given state $s$ and any given complete action profile $\delta$ there might be several possible next states. If probability of survival is more then zero, than condition 4 above requires that there should be at least one next state. If the probability of survival is equal to zero, then there are two possibilities: (i) the system has no next state, thus it always fails, or (ii) the system has one or more next states that the system might transition to with probability zero.

Next is the key definition of this paper. It's item 4 formally specifies the semantics of the modality $[C]_p$. In this definition we use term *action profile of a coalition* to refer to a function $\delta$ that assigns an action $\delta(a)$ to each agent $a$ of a coalition $C$. Also, note that for any two relations $R_1, R_2 \subseteq X \times Y$, we have $R_1 \subseteq R_2$ if every pair $(x, y) \in X \times Y$ in relation $R_1$ is also in relation $R_2$. If $f$ and $g$ are partial functions (functional relations), then $f \subseteq g$ means that function $g$ is an extension of function $f$.

**Definition 3** *For any state $s \in S$ of a transition system with catastrophic failures $(S, D, P, M, \pi)$ and any formula $\varphi \in \Phi$, satisfiability relation $s \Vdash \varphi$ is defined recursively:*

1. *$s \Vdash p$ if $s \in \pi(p)$,*
2. *$s \Vdash \neg\varphi$ if $s \nVdash \varphi$,*
3. *$s \Vdash \varphi \to \psi$ if $s \nVdash \varphi$ or $s \Vdash \psi$,*
4. *$s \Vdash [C]_p\varphi$ when there is an action profile $\delta \in D^C$ of coalition $C$ such that for any complete action profile $\delta' \in D^A$ if $\delta \subseteq \delta'$, then*
   *(a) $P(s, \delta') \ge p$ and*
   *(b) for any state $s' \in S$, if $(s, \delta', s') \in M$, then $s' \Vdash \varphi$.*

## Logical System

In addition to propositional tautologies in language $\Phi$, our logical system contains the following axioms:

1. Cooperation:
   $[C_1]_p(\varphi \to \psi) \to ([C_2]_q\varphi \to [C_1 \cup C_2]_{\max\{p,q\}}\psi)$,
   where $C_1 \cap C_2 = \varnothing$,

2. Monotonicity: $[C]_p\varphi \to [C]_q\varphi$, where $q \le p$,

3. Unachievability of Falsehood: $\neg[C]_p\bot$, where $p > 0$.

The Cooperation axiom in the form without subscripts goes back to (Pauly 2001; 2002). Informally, it says that two coalitions can combine their strategies to achieve a common goal. The assumption that coalitions $C_1$ and $C_2$ are disjoint is important because a hypothetical common agent of these two coalitions might be required to choose different actions under strategies of these two coalitions. Our version of the Cooperation axiom adds probability of survival subscript to the original version of this axiom. Perhaps one might think that the conclusion of the axiom should have subscript $\min\{p, q\}$ rather than $\max\{p, q\}$. This is not true because, according to Definition 3, statement $[C]_p\varphi$ means that coalition $C$ has a strategy to achieve $\varphi$ with probability of success $p$ *regardless* of what actions are chosen by the other agents. The Monotonicity axiom says that if a coalition $C$ can achieve goal $\varphi$ with probability at least $p$, then coalition $C$ can achieve $\varphi$ with probability $q$, where $q \le p$. Finally, the Unachievability of Falsehood axiom says that no coalition can achieve falsehood with a positive probability.

We write $\vdash \varphi$ if formula $\varphi$ is provable from the above axioms using the Modus Ponens, the Necessitation, and the Monotonicity inference rules:

$$\frac{\varphi, \varphi \to \psi}{\psi} \qquad \frac{\varphi}{[C]_0\varphi} \qquad \frac{\varphi \to \psi}{[C]_p\varphi \to [C]_p\psi}.$$

Notice that the Necessitation inference rule with positive subscript is not, generally speaking, valid. Indeed, formula $\top$ is universally true but coalition $C$ may not have a strategy that guarantees the survival of the system with a positive probability. Thus, $[C]_p\top$ is not a universally true formula for $p > 0$.

We write $X \vdash \varphi$ if formula $\varphi$ is provable from the theorems of our logical system and a set of additional axioms $X$ using only the Modus Ponens inference rule.

**Theorem 1 (soundness)** *If $\vdash \varphi$, then $s \Vdash \varphi$ for each state $s \in S$ of each system with catastrophic failures $(S, D, P, M, \pi)$.* $\boxtimes$

**Theorem 2 (completeness)** *If $X \nvdash \varphi$, then there is a state $s$ of a transition system with catastrophic failures such that $s \Vdash \chi$ for each $\chi \in X$ and $s \nVdash \varphi$.* $\boxtimes$

## References

Pauly, M. 2001. *Logic for Social Software*. Ph.D. Dissertation, Institute for Logic, Language, and Computation.

Pauly, M. 2002. A modal logic for coalitional power in games. *Journal of Logic and Computation* 12(1):149–166.