games

MDPI

Article

Game Theory Narratives of the Three-Body Trilogy

Noemi Navarro [1] and Jean-Christophe Pereau [2,*]

[1] University of Rennes, CNRS (CREM, UMR 6211), F-35000 Rennes, France; navarro.prada@gmail.com
[2] University of Bordeaux, CNRS, INRAE (BSE, UMR 6060, UMR 1441), 33600 Pessac, France
* Correspondence: jean-christophe.pereau@u-bordeaux.fr

Abstract

This article offers a formal analysis of the paradoxes, dilemmas and strategic interactions explored in Liu Cixin's trilogy *The Three-Body Problem*. Several games, such as the survival game, the deterrence game, the first contact game, and the big bang game, provide the foundations of cosmic sociology.

Keywords: cosmic sociology; analytic narratives; game theory; Dark Forest; Trisolaris

1. Introduction

Literature and game theory have a reciprocal relationship: literature has inspired game theorists, while game theory has provided new lenses for literary analysis. Von Neumann and Morgenstern (1944) drew upon Arthur Conan Doyle's Sherlock Holmes adventures to illustrate strategic reasoning in a zero-sum two-person game (Brams, 1994). Conversely, game-theoretic frameworks have been applied to ancient biblical texts by Brams (2003, 2011) and to Jane Austen's novels by Chwe (2013). Game theory models have also been used in Analytic Narratives to address outstanding puzzles in the historical literature (Bates et al., 1998; Greif, 2002). It has been shown, for instance, that Napoleon at Waterloo adopted a cautious strategy even though the battle was lost (Mongin, 2017), and that the suicide thesis is not needed to explain Caesar's death (Crettez & Deloche, 2018).

According to Suvin (1979, p. 7), science fiction is a cognitive genre, unlike mythic texts such as folktales and religious works, which conceive human relations as fixed and supernaturally determined. Science fiction provides an analytical framework for exploring logical consequences and rational human behaviors—much as game theory does. Beletto (2009) shows how the narrative, popularized in print media, that game theory was America's secret weapon during the Cold War, inspired Philip K. Dick's novel *Solar Lottery* (Dick, 1955), which imagines a political system based on the Minimax concept. No formal analysis is provided in that novel. This article aims to extend this tradition by examining Liu[1] Cixin's science fiction trilogy[2] through the lens of game theory, exploring how the novels' fictional discipline of cosmic sociology[3] can be interpreted through game-theoretic concepts.

For readers unfamiliar with the trilogy: Liu Cixin narrates the tale of two civilizations, Earth and Trisolaris, from their first contact to the end of life in the universe. The narrative is built around three main characters. The first is Professor Ye Wenjie, who reveals Earth's position to Trisolaris and offers Luo Ji, as a gesture of atonement, an intellectual weapon to defeat Trisolaris: the theoretical foundations of cosmic sociology. The second is Luo Ji, whose name in Chinese means "logic", and who plays the roles of Wallfacer and Swordholder. The third is Cheng Xin, another scientist, who becomes the second Swordholder and upon whom the destiny of the universe depends.

Academic Editor: Kjell Hausken

Received: 15 February 2026
Revised: 16 April 2026
Accepted: 22 April 2026
Published: 27 April 2026

Cosmic sociology is the study of the emergent properties of a supersociety formed by the vast number of civilizations distributed throughout the universe. These civilizations, as numerous as the detectable stars, form the body of a cosmic society. *Like Euclidean geometry, you'll set up a few simple axioms at first, then derive an overall theoretic system using those axioms as a foundation*. Cosmic sociology rests on two axioms. According to the first one: *survival is the primary need of civilization*. The second axiom states that *civilizations continuously grow and expand, but the total matter in the universe remains constant. The universe is big, but life is bigger! That's what the second axiom means. The amount of matter in the universe remains constant, but life grows exponentially* (Liu, 2015, *Dark Forest*, p. 12). To paraphrase Kenneth E. Boulding's famous dictum: infinite growth is impossible in a finite universe.

The precarious existence of Trisolaran civilization, repeatedly interrupted by collapses and recoveries due to its unstable three-sun system, clearly demonstrates that survival is the conscious objective that overrides all others and influences all actions when faced with choices.

Liu Cixin's trilogy has been the subject of several papers. Dyson (2019) provides a political science analysis through the lens of Kenneth N. Waltz's images of international relations, arguing that human nature is largely malevolent and that Trisolaris is not a morally advanced civilization with beneficent ideas about pan-species harmony. However, we argue in what follows that the Waltzian zero-sum thinking is not relevant enough to account for the paradoxes and dilemmas of the trilogy. Yu (2015) draws an interesting connection between the Dark Forest rule and Hobbes's state of nature, using the prisoner's dilemma as an analytical framework but without proposing an explicit model.

This article uses game theory to analyze the trilogy's key strategic interactions within a formal analytical framework. It describes the Dark Forest, explaining why a world-killer will always emerge when an advanced civilization is detected—showing that the Dark Forest itself is the source of universal fear, and that without it, cooperation between Earth and Trisolaris might have been possible. While remaining hidden is the best survival strategy, we show why Professor Ye Wenjie's decision to reveal Earth's coordinates to Trisolaris is optimal given her beliefs. Similarly, Trisolaris' beliefs about the type of the Swordholder explain the end of the deterrence era. Our methodology for translating the literary narrative into a formal representation is as follows: for each game, we use quotes from the trilogy to justify the ordinal ranking of each character's preferences across the scenarios they face, from most to least favorable, then convert this ranking into numerical values.

The outline of the paper follows the chronology of the trilogy by presenting several games, such as the survival game, the benevolent–malice game, the first contact game, the deterrence game, and the new big-bang game. We will aim to analyze them in the simplest way possible.

2. The Survival Game

> *The universe is a Dark Forest. Every civilization is an armed hunter stalking through the trees like a ghost. The hunter has to be careful, because everywhere in the forest are stealthy hunters like him. If he finds other life, there's only one thing he can do: open fire and eliminate them. In this forest, hell is other people. An eternal threat that any life that exposes its own existence will be swiftly wiped out. This is the picture of cosmic civilization*. (Liu, 2015, *Dark Forest*, p. 469)

The survival game illustrates the universal struggle for survival among civilizations. Consider two known civilizations in the universe: Earth and Trisolaris, both aware that other unknown civilizations may exist. Is it in their interest to reveal their existence to one another? Each civilization has two strategies: "reveal its existence" (denoted by E) and "stay hidden" (denoted by H). Both players act simultaneously. The survival game is summarized in the payoff bimatrix (Table 1). A payoff of 0 corresponds to the planet's destruction, which is the worst outcome for both, while 1 represents their survival. In each cell, the first number gives Earth's payoff, and the second one gives Trisolaris' payoff. This bimatrix shows that if a civilization reveals its position, it invites destruction by the other civilization (cells EH for Earth and HE for Trisolaris) or by an unknown one (cell EE). The best outcome occurs when both civilizations remain hidden (cell HH).

The payoff bimatrix of the survival game is

Table 1. The survival game: (H, H) is the equilibrium.

Earth/Trisolaris	E	H
E	0, 0	0, 1
H	1, 0	1, 1

It follows directly that for both Earth and Trisolaris, remaining hidden is a strictly dominant strategy. Regardless of the opponent's choice, hiding is always preferable. Since the game is symmetric, both civilizations will adopt this strategy to ensure survival. The equilibrium is therefore (H, H) with a payoff of (1, 1).

3. But Who Is Going to Attack?

Even if Earth and Trisolaris mutually decide to reveal their existence, at least one civilization will choose to attack because, for one technologically advanced species, shooting a bullet may be less costly than exploration. Revealing one's position becomes a fatal curse. Dark Forest attacks share two qualities: they are *casual* and *economical*. Casual means that *the only basis for the attack is the exposure of the target's location. There will be no reconnaissance or exploration conducted against the target beforehand. For a supercivilization, such exploration is more expensive than a blind strike. Economical means that the attack will employ the least expensive method: using a small, worthless projectile to trigger the destructive potential already present in the target star system* (Liu, 2016, *Death's End*, p. 240).

Could an anonymous attack be prevented simply because each civilization waits for another to act first? This behavior of free riding, or diffusion of responsibility, was proposed by Darley and Latane (1968) and modeled by Diekmann (1985) under the name of the volunteer dilemma game[4]. In our case the volunteer is a world-killer.

This game concerns the millions of civilizations in the Dark Forest (estimated at 1.57 million). Faced with the discovery of planet Earth, the question is whether any civilization will launch an attack against it. We take the perspective of one civilization facing N others. Each civilization chooses between launching the photoid (denoted L) or waiting for another to do so (denoted W). The civilization that launches the photoid incurs a cost c, which may be very low but is strictly positive. Once Earth's destruction is confirmed, all civilizations benefit from the amount of matter returned to the universe. This positive gain, denoted v, is a public good: it benefits everyone freely (non-rival) and no civilization can be excluded from enjoying it (non-excludable). We assume $v > c$.

Let us consider the dilemma between two civilizations ($N = 2$), namely Sophia and Electra, to determine which will be the world-killer. The payoff bimatrix of the game is shown in Table 2.

Table 2. The world-killer game in normal form.

Sophia\Electra	W	L
W	0, 0	$v, v - c$
L	$v - c, v$	$v - c, v - c$

This game admits two asymmetric equilibria in pure strategies: (W, L) and (L, W). To determine the symmetric equilibrium in mixed strategies, we denote the probability of choosing W by p (L is then chosen with probability $1 - p$). In a symmetric equilibrium, the probability of choosing W is the same for both players. Sophia's expected payoff is equal to $EU_{Sophia}(W) = p \times 0 + (1 - p).v$ when she chooses W, and to $EU_{Sophia}(L) = p \times (v - c) + (1 - p) \times (v - c) = v - c$ when she chooses L. Sophia must be indifferent between W and L in order to play a mixed strategy. This will happen when $EU_{Sophia}(W) = EU_{Sophia}(L)$ yielding $p^* = c/v$ as the (mixed-strategy) equilibrium probability of choosing W. The equilibrium probability of playing W increases with the individual cost and decreases with the benefit. The situation is symmetric for Electra.

To determine the symmetric equilibrium in mixed strategies with N players, we again denote by p the probability of choosing W for all players (symmetric equilibrium). Sophia's expected gain from playing W and L is now equal to $EU_{Sophia}(W) = (1 - p^{N-1})v$ and $EU_{Sophia}(L) = v - c$, respectively. This is so because with probability p^{N-1} none of the other players wait, which means that with probability $1 - p^{N-1}$ at least one of the other players launches the photoid. Sophia will be indifferent between the two actions for $p^* = (v/c)^{1/(N-1)}$. At equilibrium, the probability of all N players waiting is equal to $(p^*)^N = (c/v)^{N/(N-1)}$ and the probability of at least one player launching the protoid is equal to $1 - (p^*)^N$. As N grows larger, $1 - (p^*)^N$ approaches $1 - c/v$, which is not zero. Suppose that c is equal to 10% of the gain v, which gives a ratio of $c/v = 0.1$. For $N = 2$, the probability of attacking is $1 - (p^*)^N = 0.99$; for $N = 20$ it is $1 - (p^*)^N = 0.91$; but for very large N, $1 - (p^*)^N$ approaches 0.9 and not 0. If the cost is very high, around 90% of the gain, the probability will be lower, 0.1, but not zero. As pointed out by Yu (2015), *if the universe is viewed as a closed system, entropy increase is irreversible and other civilizations must be eliminated to slow down the heat death of the universe*.[5] The attack will definitely occur.

In this respect, the most significant difference between the two civilizations was that Trisolaris had long known that the universe was a Dark Forest, unlike Earth. Trisolaris even found it strange that Earth had taken so long to understand this.

4. To Be or Not to Be . . . Benevolent

Let us assume now that the universe contains only two civilizations, Earth and Trisolaris, each knowing the other one's existence and location. In this situation, could the two civilizations at least coexist without attacking each other?

To show this, we assume that each civilization can be *benevolent* (strategy B) or *malicious* (strategy M). *Benevolence means refraining from initiating an attack to eradicate the other civilization; malice represents the opposite intent* (Liu, 2015, *Dark Forest*, p. 465). If both civilizations are benevolent, they ensure their survival, each having a payoff equal to 1 (cell BB in Table 3). On the other hand, a benevolent civilization can be annihilated by a malicious one (cells BM and MB). The worst-case scenario is mutual destruction if they both adopt the strategy M (cell MM). As in the survival game, 0 means annihilation of the planet and 1 survival.

This scenario can be summarized in the following payoff bimatrix:

Table 3. The benevolent–malice game: all outcomes are Nash equilibria.

Earth/Trisolaris	B	M
B	1, 1	0, 1
M	1, 0	0, 0

One might think that strategies B and M are payoff-equivalent for each civilization, but the outcomes differ significantly. If both civilizations are benevolent, neither will attack, and both will survive (a payoff of 1 for each). If both are malicious, they will mutually destroy each other (0 for each). If one is benevolent while the other one is not, the benevolent civilization seals its own fate. Clearly, mutual benevolence is preferable to the extinction of all civilizations in the universe. Each civilization faces a social dilemma. The outcome (B, B) is superior to (M, M) for both civilizations, but nothing guarantees it will be chosen. In fact, this game admits multiple Nash equilibria[6]. If Earth plays B, Trisolaris' best response can be either B or M (it survives either way). Suppose Trisolaris chooses M; Earth's best response is then B or M (it will be destroyed regardless). If Earth chooses B, the outcome (B, M) constitutes a Nash equilibrium. All four outcomes are, in fact, Nash equilibria. We thus face a multiplicity of equilibria and therefore indeterminacy. The reason lies in the fact that each civilization's payoff depends solely on the other one's decision. Whatever its choice, everything hinges on its opponent's action. Since each outcome is an equilibrium, it is individually rational for Earth to adopt benevolence even when Trisolaris decides to be malicious, yielding the worst possible configuration for Earth with the outcome (B, M).

As pointed out by Luce and Raiffa (1957, p. 107), in games where every pair of pure or mixed strategies is an equilibrium, we can use the notion of domination to refine the set of equilibria. A pair of strategies *jointly dominates* another one if both players strictly prefer the former to the latter. A *jointly admissible* strategy is a pair of strategies that is not jointly dominated by another one. In this case, the outcome of mutual malevolence (M, M) is jointly dominated by the outcome of mutual benevolence (B, B). Furthermore, the outcome of mutual benevolence *weakly* jointly dominates (M, B) and (B, M), meaning that no player is worse off in (B, B) than in (M, B) or (B, M). This refinement allows us to select the outcome of mutual benevolence (B, B), showing that if it were not for the Dark Forest, cooperation between Earth and Trisolaris might have been possible. If we abandon the assumption of only two civilizations in the universe, we are back to the first game.

Liu Cixin develops the concept of the chain of suspicion: *If you think I'm benevolent, that's not a reason to feel safe, because according to the first axiom, a benevolent civilization can't predict that any other civilization is benevolent. You don't know whether I think you're benevolent or malicious. Next, even if you know that I think you're benevolent, and I also know that you think I'm benevolent, I don't know what you think about what I think about what you're thinking about me. It's convoluted, isn't it? This is just the third level, but the logic goes on indefinitely* (Liu, 2015, *Dark Forest*, p. 466). This chain of suspicion refers to the assumption of common knowledge of rationality which can be considered as the most restrictive principle in game theory. While Schelling (1960) and Lewis (1969) first suggested this hypothesis, it was Aumann (1976) who provided an analytical definition (Brandenburger & Dekel, 1991). The common knowledge of rationality hypothesis assumes an infinite and complex game of anticipation that allows actors to select equilibrium strategies. All reasoning takes place at equilibrium without any reference to what might have happened before the equilibrium situation was reached. Schelling (1960) rightly illustrates the cognitive complexity introduced by this common knowledge of rationality hypothesis. He gives a revealing example: a man loses his wife in a department store; will they be able to find each other again? The problem is not simply knowing what the other person will do; it is not a problem of objective prediction. What the other person will do depends on how the second player predicts the first player's action, knowing that the first player is also trying to put himself in the second player's position, and so on. According to Schelling, in practice, this problem can be solved: the players manage to coordinate since they know that the other player is also trying to do so. Players can therefore find a solution in this type of game when they share a minimum of cultural similarities in an interconnected ecosystem. This is obviously not the case for Earth and Trisolaris. Earth thought that the entire galaxy was an empty desert.

However, in the first two games, we assumed that both civilizations have the means to launch a tiny object traveling close to the speed of light (call it a photoid) and that it would be capable of destroying a planet. This is not the case for Earth. Furthermore, destroying Earth is not Trisolaris' goal, as Trisolaris wants to settle on planet Earth, not to destroy it. This suggests considering new games with asymmetric preferences.

Trisolaris' position is expressed by its chancellor who said: *We cannot share the Earth with the people of that world. We could only destroy Earth civilization and completely take over that solar system. But there is another reason for destroying Earth civilization. They're also a warlike race. Very dangerous. If we try to coexist with them on the same planet, they will shortly learn our technology. Continuing in that state would allow neither civilization to thrive* (Liu, 2014, *Three-Body*, p. 287).

In terms of preferences, this means that the situation of mutual benevolence (B, B) is not the best outcome for Trisolaris, which prefers to be malicious when Earth is benevolent (B, M). Moreover, since Trisolaris has a technological advantage, it believes that it will win the war if Earth is malicious. The worst situation for Trisolaris is to be benevolent while Earth is malicious. From Earth's perspective, mutual benevolence is the best outcome. Furthermore, Earth prefers being benevolent to being malicious with a benevolent Trisolaris. But once Trisolaris is malicious, the human race is doomed to disappear. Earth's preference ranking over the cells in the bimatrix in Table 4 below (each cell representing a combination of players' strategies, Earth's strategy noted first) is BB = 3 > MB = 2 > BM = MM = 0 while Trisolaris' preference ranking is BM = 2 > BB = MM = 1 > MB = 0. These preferences show that malice is a dominant strategy for Trisolaris, and whatever Earth does, even if she prefers to be benevolent to malicious, the human race will be annihilated in both Nash equilibria (B, M) and (M, M) of the game.

Table 4. The modified benevolent–malice game: (B, M) and (M, M) are Nash equilibria.

Earth/Trisolaris	B	M
B	3, 1	0, 2
M	2, 0	0, 1

In the Dark Forest, mutual cooperation cannot exist. The next game tackles the crucial issue of the first contact between the two civilizations.

5. The First-Contact Game

> *This world has received your message. I am a pacifist in this world. It is the luck of your civilization that I am the first to receive your message. I am warning you: Do not answer! Do not answer!! Do not answer!!!* To this message sent by the listener at Post 1379 on Trisolaris, Professor Ye response was *Come here! I will help you conquer this world. Our civilization is no longer capable of solving its own problems. We need your force to intervene.* (Liu, 2014, *Three-Body*, p. 222)

This is a key moment in the trilogy. The first message Professor Ye received begged her not to respond. Hence, why did she decide to reveal Earth's coordinates to Trisolaris? This is because she had learned of Trisolaris' existence, its cycles of collapse and rebirth, and its plan to migrate to another star system.

In order to understand her decision, we combine the strategy spaces of the two previous games. Earth can reveal its presence or remain hidden while Trisolaris can be benevolent or malicious. The best outcome for Earth is to reveal its presence when Trisolaris is benevolent while the worst outcome occurs when Trisolaris is malicious. In Table 5, Earth's preference ranking over each combination of strategies is EB = 2 > HB = HM = 1 > EM = 0 while Trisolaris' prefers to be malicious rather than benevolent when Earth reveals its existence: EM = 3 > EB = 2 > HB = HM = 1.

Table 5. The first-contact game: (H, M) is a Nash equilibrium.

Earth/Trisolaris	B	M
E	2, 2	0, 3
H	1, 1	1, 1

The outcome (H, M) is a Nash equilibrium. If we consider that the universe contains more than two civilizations, Earth faces the risk of being attacked by a third civilization, leading to a payoff of 0 when she plays E and Trisolaris B. In that case, to stay hidden for Earth becomes a strictly dominant strategy.

The normal form game in Table 5 corresponds to the situation when each civilization's moves are simultaneous or unobservable by the other civilization. If at least one civilization's moves are sequential and observable by the other, then the situation corresponds to a sequential game (with perfect information). Earth moves first and decides whether to reveal its presence or not, then Trisolaris can decide whether to be benevolent or not after observing Earth's choice. This game in extensive form is represented in the tree depicted in Figure 1. In the payoff pairs noted after each of the terminal branches in Figure 1, Earth's payoff appears in the first coordinate while Trisolaris' payoff appears in the second coordinate.

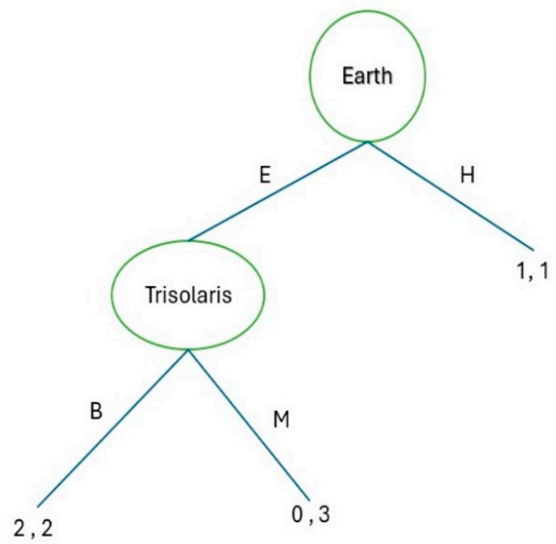

Figure 1. Extensive-form of the first contact game.

The game is solved by backward induction. As the best strategy of Trisolaris is to play M (because 3 > 2), the best response of Earth, anticipating Trisolaris choice, is to remain hidden (as 1 > 0). We can relax the assumption of a universe made up of two civilizations. As before, even if we assume that a benevolent Trisolaris will not attack Earth, another civilization can do it, implying in all cases the destruction of the Earth (with a payoff equal to 0 in the issue EB). Earth's best strategy is to stay hidden.

So, how can we explain the decision of Professor Ye? We need to assume that she did not know Trisolaris' intentions concerning Earth. This means that the game between Earth and Trisolaris is one of incomplete information, also known as a Bayesian game.

6. The Bayesian First-Contact Game

Let us assume that Nature determines the type of Trisolaris. As shown in Figure 2, with a probability of $0 < p < 1$, Trisolaris is benevolent, and with a complementary probability $(1 - p)$, it is malicious. Trisolaris knows its type (this is its private information), but Professor Ye does not. Trisolaris has the choice between two actions: attack A or cooperate C, while Earth can reveal its existence E or remain hidden H. The set of strategies for Earth is therefore $S_{Earth} = \{E, H\}$, while Trisolaris has four strategies $S_{Trisolaris} = \{AA, AC, CA, CC\}$. The strategy AA means playing A when Trisolaris is of type B and also A when Trisolaris is of type M. The strategy AC means playing A when Trisolaris is of type B and playing C when Trisolaris is of type M. When Trisolaris is of type M, we have the same payoff structure as in Figure 1. When Trisolaris is of type B, not attacking gives Trisolaris a higher gain than attacking after Earth revealed its position. Trisolaris gets 2 if the outcome is (E, C) for both types, but if the outcome is (E, A), Type-B Trisolaris gets 0 while Type-M Trisolaris gets 3. Indeed, Type-B Trisolaris prefers not being aware that a civilization exists rather than knowing about it and having to destroy it. This is the only difference in terms of preferences between type B and type M. On the other hand, Type-M Trisolaris prefers the (E, A) outcome to (E, C) (because 3 is higher than 2). When Earth remains hidden, its gain is 1 independent of Trisolaris's type. Whatever the type of Trisolaris, Earth's preference ranking is the same EC = 2 > HA = HC = 1 > EA = 0. Trisolaris' preference ranking when of type-M is EA = 3 > EC = 2 > HA = HC = 1 while Trisolaris' preference ranking when of type B is EC = 2 > HA = HC = 1 > EA = 0. EA, which was Trisolaris' best combination of strategies when of type M, becomes the worst choice when of type B.

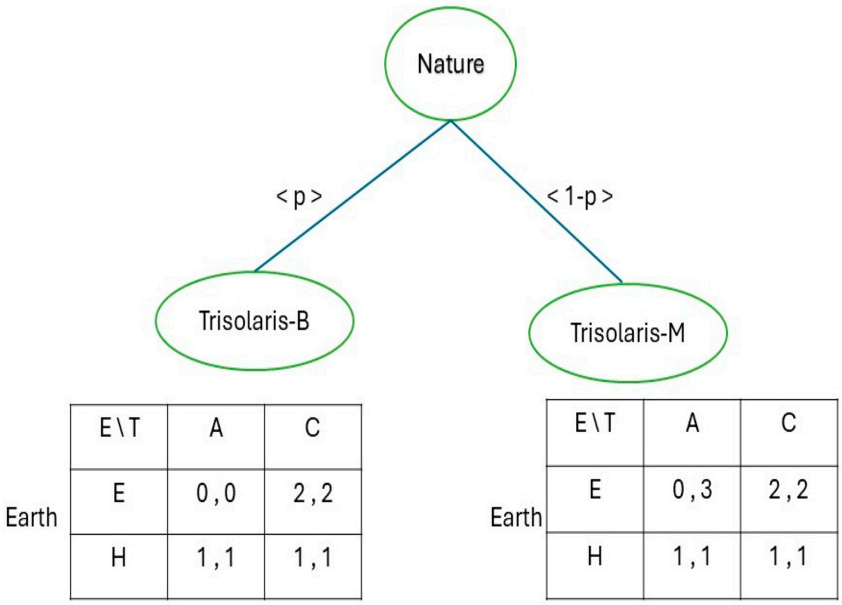

Figure 2. Extensive-form of the Bayesian first-contact game.

From this, we derive the following normal-form payoff bimatrix (Table 6):

Table 6. Normal form of the first-contact game.

Earth/Trisolaris	AA	AC	CA	CC
E	$0, 3 \times (1 - p)$	$2 \times (1 - p), 2 \times (1 - p)$	$2 \times p, 3 - p$	2, 2
H	1, 1	1, 1	1, 1	1, 1

There are several Bayes–Nash equilibria. For example, the outcome of the game (E, CA) leading to the payoff (2, 2) is one of them as shown in Table 6. For this equilibrium to hold, condition $2p \geq 1$ must be verified, i.e., $p \geq 1/2$ for Earth, and conditions $3 - p \geq 3 \times (1 - p)$, $3 - p \geq 2 \times (1 - p)$, and $3 - p \geq 2$ must be verified for Trisolaris. The outcome of the game (E, CC) is also a Bayes–Nash equilibrium, but only for $p = 1$. Thus, if Professor Ye's belief that Trisolaris is benevolent is greater than or equal to 0.5, then her optimal strategy is to reveal Earth's coordinates.

Professor Ye will learn too late that Trisolaris' intentions were not in line with her original ideal of using the alien civilization as a way to reform humanity. After learning the truth about Trisolaris's civilization, Professor Ye becomes silent. The reality of Trisolaris's civilization was different from her *beautiful fantasies*. She believed that *a society with such advanced science must also have more advanced moral standards* (Liu, 2014, *Three-Body*, p. 281).

Professor Ye's decision was not the right choice for Earth because Trisolaris' true type was malicious. Unlike Trisolaris, Earth did not know that the universe was a Dark Forest. By launching its fleet across space to find and occupy a new planet, Trisolaris's decision became irreversible—retreat was impossible, as illustrated by the idea of burning its own ships[7]. To prevent Earth from communicating its position to any other civilization and to block any major technological improvement, Trisolaris placed Earth under quarantine.

7. The Deterrence Game

Is the end of humanity coming? The Wallfacer Luo Ji manages to make the following threat. Using a mechanism of bombs deployed throughout the solar system, Luo Ji threatens to detonate them to reveal the relative positions of both Trisolaris and Earth, which would inevitably lead to the annihilation of both planets. The threat was credible enough for Trisolaris proposing to negotiate, but Luo Ji replies with a take-it-or-leave-it offer: refrain from destroying human civilization and cooperate. The threat allows both civilizations to guarantee their survival. However, one question remains about the credibility of carrying out this threat, which leads us to the game of deterrence.

This new game corresponds to a sequential version of the one between the deterrer (Earth's representative in the person of the Swordholder) and the deterred party (Trisolaris). The extensive-form game is shown in Figure 3, with Trisolaris now moving first and Earth second (Trisolaris payoffs are noted in the first coordinate at each payoff pair).

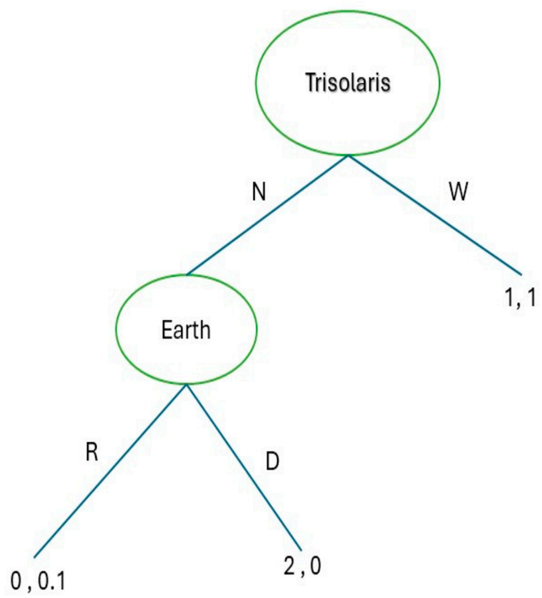

Figure 3. Extensive-form of the deterrence game.

Trisolaris has the choice of either neutralizing Earth's defense mechanism, which involves revealing its coordinates in the Dark Forest (N), or waiting (W). Earth has the choice between revealing its coordinates (R), which will lead to the annihilation of both civilizations (a death strategy), or not reacting (D), which will lead to its own destruction. In Figure 3, the status quo corresponds to a payoff of 1 for both players. The best outcome for Trisolaris is to attack with no reaction of Earth. However, if Earth reveals both their coordinates, they will both disappear. Nevertheless, Earth prefers punishing Trisolaris to end the deterrent game (NR) instead of ending up in slavery (ND), i.e., 0.1 > 0. Anticipating that the choice of Earth will be R, Trisolaris will play W. Earth's preference ranking is (following the order of play) W = 1 > NR = 0.1 > ND = 0, while Trisolaris' preference ranking is ND = 2 > W = 1 > NR = 0 for Trisolaris.

The threat to reveal Trisolaris' position is considered credible by Trisolaris, who prefers to wait. If Trisolaris plays N, then Earth will choose R, which will lead Trisolaris to play W. Thus, Trisolaris faces a choice: wait and survive by not attacking, or attack with the risk of destruction. This was the situation Trisolaris faced when Luo Ji became Earth's first Swordholder, the one responsible for carrying out the threat, after saving Earth as a Wallfacer.

The question we address now is why Trisolaris decided to end the deterrence era by launching six droplets at the speed of 25,000 km/s in Earth's direction. For that we need to take into account the beliefs of Trisolaris on the type of Swordholder in charge of carrying out the threat.

Let us assume that Nature determines the type of the Swordholder. As shown in Figure 4, with a probability $0 < p < 1$, the Swordholder is a Strong (S) type and with the complementary probability she is a Weak (W) type. When facing a Strong (S) type, Trisolaris's strategy is to wait but face against a Weak (W) type; neutralizing Earth's defense system becomes a winning strategy. The payoffs with a type-S Swordholder are the same as before (see Figure 3). The only difference lies in the payoff of the Swordholder when he/she plays R. A type W prefers not reacting to an attack from Trisolaris while a type S prefers to counterattack. The preference ranking for a type-W Swordholder becomes W = 1 > ND = 0 > NR = −0.1 while it remains the same for Trisolaris.

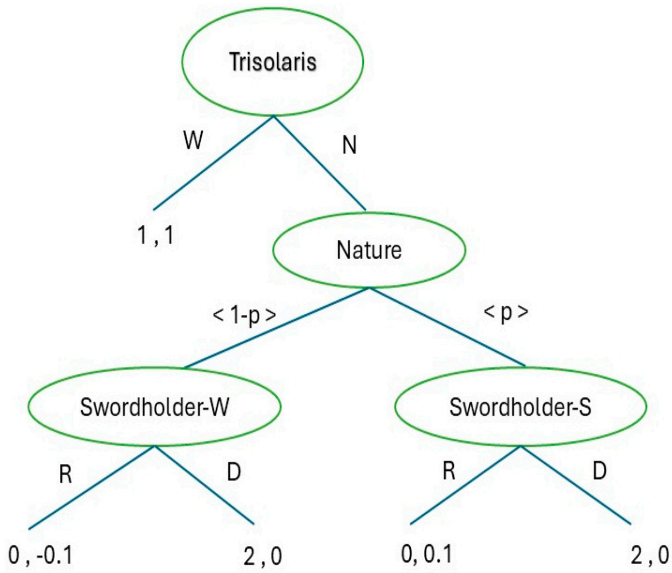

Figure 4. Extensive-form of the modified deterrence game.

With respect to the previous game: if the Swordholder's type is weak, he or she will choose D, since 0 > −0.1. If the type is strong, he or she will choose R. Trisolaris's expected payoff from playing N is 2(1 − p) + 0·p, while its payoff from playing W is 1. It follows that for p < 1/2, Trisolaris will choose to end the deterrence game.

This is precisely what happened when Earth replaced Luo Ji with Cheng Xin as Swordholder: Trisolaris seized the initiative and attacked the infrastructure on which the threat depended. The new Swordholder had fifteen minutes to react and broadcast the coordinates of both civilizations—a signal that could trigger mutual destruction. She could not bring herself to make that choice, even though humanity's end was already sealed by Trisolaris's attack. *Four billion years weighed upon her heart and left her paralyzed. Because the universe is not a fairy tale, and Trisolaris deemed her to be one. In her subconscious, she was a protector, not a destroyer; she was a woman, not a warrior. Pressing the switch would end the progress of 3.5 billion years. Everything would disappear in the eternal night of the universe, as though none of it had ever existed* (Liu, 2016, *Death's End*, p. 154).

For deterrence to hold, the probability p had to be sufficiently high. Luo Ji's probability was assessed to be between 91.9% and 98.4%—he had already proven himself capable of destroying a world. At that level, Trisolaris could not afford to gamble. With Cheng Xin, the probability was below 10%.

This marked the end of the Dark Forest deterrence. Deterrence, in this context, refers to Earth's neutralization of Trisolaris's belligerent intentions—not an equilibrium of compromise.

If the collapse of deterrence is likely to cause one or both civilizations to disappear, the prospect of the universe's end raises an altogether different game.

8. The Big-Bang Game

> *The total mass of our universe has decreased to below the critical threshold. The universe will turn from being closed to open, and die a slow death in perpetual expansion. All lives and all memories will also die. Please return the mass you have taken away and send only memories to the new universe.* (Liu, 2016, *Death's End*, p. 518)

This is the message propagated by the Zero-Homers (also called the Resetters) across the universe. As the universe is dying, the Zero-Homers want to reset it and return it to its original state, the Garden of Eden. However, the Trisolaran civilization, along with numerous other civilizations, has constructed several hundred mini-universes, each of which has taken away some matter from the great universe.

This message creates the conditions for a coordination problem. Each mini-universe faces two options: return its mass—thereby ensuring the great universe collapses into a singularity and triggers a new Big Bang—or refuse to do so. Refusal implies the death of the great universe, but *people can live out the rest of their lives in this mini-universe*. As Cheng Xin (the second Swordholder, living in Universe 647) observes, *if everyone in every mini-universe thinks that way, then they will have doomed the great universe* (Liu, 2016, *Death's End*, p. 519).

In terms of preferences, if the mini-universes fail to cooperate, their situation remains unchanged (the DD cell, with a payoff of 1 for each in Table 7); if they cooperate, their payoff is higher (the CC cell, with a payoff of 2 for each). The worst-case scenario occurs when one cooperates and the other one does not. The preference ranking for Universe 647 is CC = 2 > DC = DD = 1 > CD = 0, and symmetrically for the RoU (Rest of Universe). The new Big Bang game can thus be written as

Table 7. The new big-bang game: (C, C) and (D, D) are Nash equilibria.

Universe 647/RoU	C	D
C	2, 2	0, 1
D	1, 0	1, 1

This game has two Nash equilibria, (C, C) and (D, D), which are Pareto-ranked: the first one yields a better outcome for all players than the second one.

It is also possible to derive the mixed-strategy equilibrium. Suppose Universe 647 plays C with probability $0 < p < 1$ and D with the complementary probability $(1 - p)$, while RoU plays C with probability $0 < q < 1$ and D with probability $(1 - q)$. Universe 647's expected payoff from playing C is $EU_{U647}(C) = 2q$, and its expected payoff from playing D is $EU_{U647}(D) = q \cdot 1 + (1 - q) \cdot 1 = 1$. Universe 647 is indifferent between the two strategies when $EU_{647}(C) = EU_{647}(D)$, which yields $q = 1/2$. By symmetry, $EU_{ROU}(C) = EU_{ROU}(D)$ holds at $p = 1/2$. In the mixed-strategy equilibrium, each player assigns equal probability to C and D.

Cheng Xin ultimately decides to return matter from the mini-universe to the great universe. The outcome (C, C) is Pareto superior. *The ultimate fate of all intelligent beings has always been to become as grand as their thoughts* (Liu, 2016, *Death's End*, p. 521).

9. Conclusions

In this article we have analyzed several games inspired by Liu Cixin's *Three-Body Problem* trilogy, each illuminating a different facet of cosmic sociology. The main findings are as follows: the dominant strategy for any civilization is to remain hidden in the Dark Forest; Professor Ye's decision to reveal Earth's position is optimal based on her belief that Trisolaris was benevolent; the game of deterrence hinges on the identity of the Swordholder tasked with carrying out the threat; and the survival of the universe requires the cooperation of all mini-universes.

Rationality and emotions distinguish Trisolaris from Earth. *Trisolaris's approach to humans is a rational choice. It's the responsible thing to do for the survival of their species, and has nothing to do with good or evil* (Liu, 2015, *Dark Forest*, p. 159). *Anger, sorrow, happiness, and appreciation of beauty—were emotions that the Trisolaran civilization strove to avoid and eliminate. The history of the past two hundred-some cycles of civilization proved that civilizations that relied on calmness and numbness as their spiritual core were the most capable of survival* (Liu, 2014, *Three-Body*, p. 286). On the contrary, Cheng Xin's decisions not to commit a world-killing act or to return matter to the universe show that *humanity has love* (Liu, 2015, *Dark Forest*, p. 494).

Author Contributions: Conceptualization, N.N. and J.-C.P.; Methodology, N.N. and J.-C.P.; Formal analysis, N.N. and J.-C.P.; Writing—original draft, N.N. and J.-C.P.; Writing—review & editing, N.N. and J.-C.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: No new data were created or analyzed in this study.

Acknowledgments: We would like to thank the three reviewers for their stimulating comments.

Conflicts of Interest: The authors declare no conflict of interest.

Notes

1. Chinese names are written with surname first.
2. The trilogy comprises *The Three-Body Problem*, *The Dark Forest*, and *Death's End*. All quotations are from Ken Liu and Joel Martinsen's English translation and in italics in the text.
3. Cosmic sociology naturally refers to psychohistory created by Isaac Asimov in his Foundation cycle. Psychohistory is the branch of mathematics that deals with the reactions of human groups to constant social and economic phenomena. In 1942, Asimov founded this mathematical analysis of social structures based on the kinetic theory of gases and not on game theory. Individuals do not behave as erratically as gas molecules (or at least we hope).
4. An illustration is the murder of Kitty Genovese, killed in New York City in 1964 despite the presence of numerous neighbors who did not call the police. We now know that this interpretation is false—but unfortunately the murder itself is not.

[5] Recent advances in evolutionary game theory applied to multiplayer network games show that cooperation may appear in public good games (Wang et al., 2024; Wang & Su, 2026). It suggests future applications.

[6] Nash equilibrium is the outcome of a game where each player is using a strategy that maximizes their own expected payoff, given the strategies of the other players (Nash, 1951).

[7] Alexander the Great in his war against Persian decided to burn his vessels in 334 BC to signify to his troops that turning back was no longer possible and that the only way out was victory or death. In 1519 AD, Hernán Cortés decided to sink his ships for the same reason.

References

Aumann, R. J. (1976). Agreeing to disagree. *Annals of Statistics*, *4*, 1236–1239. [CrossRef]

Bates, R. H., Greif, A., & Levi, M. (1998). *Analytic narratives*. Princeton University Press.

Beletto, S. (2009). The game theory narrative and the myth of the national security state. *American Quarterly*, *61*(2), 333–357. [CrossRef]

Brams, S. (1994). Game theory and literature. *Games and Economic Behavior*, *6*, 32–54. [CrossRef]

Brams, S. (2003). *Biblical games: Game theory and the Hebrew bible*. Massachusetts Institute of Technology Press.

Brams, S. (2011). *Game theory and the humanities: Bridging two worlds*. The MIT Press.

Brandenburger, A., & Dekel, E. (1991). The role of common knowledge assumption in game theory. In F. Hahn (Ed.), *The economics of missing markets, information and games* (pp. 46–61). Clarendon Press.

Chwe, M. S.-Y. (2013). *Jane Austen, game theorist*. Princeton University Press.

Crettez, B., & Deloche, R. (2018). An analytic narrative of Caesar's death: Suicide or not? That is the question. *Rationality and Society*, *30*(3), 332–349. [CrossRef]

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: Diffusion of responsibility. *Journal of Personality and Social Psychology*, *8*, 377–383. [CrossRef] [PubMed]

Dick, P. K. (1955). *Solar lottery*. Ace Books.

Diekmann, A. (1985). Volunteer dilemma. *Journal of Conflict Resolution*, *29*(4), 605–610. [CrossRef]

Dyson, S. B. (2019). Images of international politics in Chinese science fiction: Liu Cixin's *three-body* problem. *New Political Science*, *41*(3), 459–475. [CrossRef]

Greif, A. (2002). Chapter 52. Economic history and game theory. In R. Aumann, & S. Hart (Eds.), *Handbook of game theory with economic applications* (Vol. 3, pp. 1989–2024). North-Holland.

Lewis, D. K. (1969). *Convention: A philosophical study*. Harvard University Press.

Liu, C. (2014). *The three-body problem*. Tom Doherty Associates, LLC, USA.

Liu, C. (2015). *The dark forest*. Tom Doherty Associates, LLC, USA.

Liu, C. (2016). *Death's end*. Tom Doherty Associates, LLC, USA.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions*. Harvard University.

Mongin, P. (2017). A game-theoretic analysis of the Waterloo campaign and some comments on the analytic narrative project. *Cliometrica*, *2*, 451–480. [CrossRef]

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics*, *54*, 286–295. [CrossRef]

Schelling, T. C. (1960). *The strategy of conflict*. Harvard University Press.

Suvin, D. (1979). *Metamorphoses of science fiction, on the poetics and history of a literary genre*. Yale University Press.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.

Wang, C., Perk, M., & Szolnoki, A. (2024). Evolutionary dynamics of any multiplayer game on regular graphs. *Nature Communication*, *15*, 5349. [CrossRef] [PubMed]

Wang, C., & Su, Q. (2026). Public goods games on any population structure. *Science Advances*, *12*, eaeb1263. [CrossRef] [PubMed]

Yu, C. (2015). The dark forest rule: One solution to the fermi paradox. *Journal of the British Interplanetary Society*, *67*, 1–3.