

Why Cooperation Occurs: Beyond the Predictions of Game Theory

Advaith Sriram
advaith.omegaz@gmail.com

ABSTRACT

Classical game-theoretic models such as the one-shot Prisoner's Dilemma predict that rational, self-interested actors will often choose not to cooperate, yet cooperation frequently occurs in real-world settings. This paper examines the factors that sustain cooperation despite the incentives that often encourage individuals to defect. Drawing on game-theoretic models, experimental studies of strategic behavior, historical case studies from World War I and organized crime, and research in behavioral psychology, this paper analyzes structural, contextual, and psychological explanations for cooperation. The paper argues that no single mechanism is sufficient to explain cooperative behavior. Repeated interaction, punishment, and reputation can sustain cooperation, but their effectiveness depends on contextual conditions such as payoff structures, observability, and the ability to respond to defection. Psychological factors, including intuition and altruistic motivation, also encourage cooperation, but they interact with—and sometimes contradict—the assumptions of traditional game-theoretic models. Together, these findings suggest that cooperation emerges through the interaction of structural mechanisms, contextual variables, and psychological processes rather than through a single universal explanation. Future research should examine how these mechanisms interact across a wider range of strategic settings, including non-Western contexts and situations involving institutional enforcement.

INTRODUCTION

Classical game-theoretic models such as the one-shot Prisoner's Dilemma predict that rational, self-interested actors will often choose not to cooperate, even if cooperation is more beneficial for everyone involved (Shepsle and Bonchek 1997). This prediction appears in a wide range of settings, including business competition, political alliances, and daily social interactions. However, despite these predictions, cooperation frequently occurs in the real world. Businesses create partnerships, countries form alliances, and individuals work together even if acting selfishly would help them more. For example, during World War I, countries that had every strategic reason to abandon their allies instead maintained costly coalitions for years (Wolford 2019). This illuminates a central dilemma within game theory: What factors explain why people cooperate in real-world strategic settings despite the theoretical prediction that rational, self-interested actors will choose not to cooperate?

Aug 2026
Vol 10, No 4.

The Prisoner's Dilemma is one of the clearest examples of this problem. In this scenario, two individuals are arrested for a crime and placed in separate interrogation rooms. They must choose whether to cooperate (stay silent, follow an already-created story) or betray (confess that the other person committed the crime). If both criminals choose to cooperate, then they will only receive a light sentence. If they both defect, then they will receive a sentence longer than if they had cooperated. However, if one cooperates and one betrays, the one that cooperates will receive the longest sentence possible, while the betrayer leaves without a sentence. As a result, both individuals end up betraying each other, even though mutual cooperation would have led to better results for both of them. This result holds even though both players know that mutual cooperation would leave them better off. The problem is not ignorance but incentives: each player's best response is to defect regardless of what the other does (Shepsle and Bonchek 1997).

However, real-world scenarios often contradict this prediction. For instance, members of the Italian Mafia often follow a code of silence known as *Omertà*, refusing to testify against one another even when offered incentives to cooperate with authorities. During the Mafia Commission Trial of the 1980s, several defendants accepted long prison sentences rather than cooperate with prosecutors (University of Virginia Law Library). Campana and Varese argue that this behavior should not be understood as voluntary cooperation alone. Instead, criminal organizations sustain cooperation through shared involvement in violent acts, which creates mutual vulnerability and credible threats of punishment for betrayal (Campana and Varese). Unlike the reciprocal cooperation described in repeated games, cooperation within criminal organizations is often maintained through coercion, organizational control, and credible commitments. Nevertheless, this example illustrates that mechanisms beyond simple self-interest can sustain cooperation. While game theory provides a clear account of why defection is rational, this paper compares structural, contextual, and psychological explanations to examine how different mechanisms sustain cooperation across strategic settings.

This paper argues that no single mechanism is sufficient to explain cooperation. Instead, cooperation emerges through the interaction of structural mechanisms such as repeated interaction and punishment, psychological factors such as intuition and altruism, and contextual variables such as payoff structure and the type of strategic setting.

METHODOLOGY

This paper uses a literature review methodology to examine why cooperation occurs despite game theory's prediction that rational actors will often defect. Sources were identified through assigned course readings and supplemented with peer-reviewed journal articles, academic books, and historical case studies relevant to cooperation in strategic interactions. Sources were selected if they examined mechanisms influencing cooperation or defection, including repeated interaction, incentive structures, contextual factors, and psychological explanations. Sources that focused primarily on mathematical proofs, technical game-theoretic models without discussion of cooperative behavior, or topics unrelated to strategic cooperation were excluded. By examining experimental evidence, theoretical models, and historical case studies, this paper demonstrates that the predictive failures of traditional game theory often

stem from treating these mechanisms in isolation rather than understanding how they reinforce or weaken one another across different strategic settings.

RESULTS

Repeated interaction fundamentally changes the incentives that shape cooperation. In one-shot interactions, individuals focus only on immediate outcomes, making defection more attractive. However, when interactions continue over time, individuals must also consider the future consequences of their actions. Dixit and Nalebuff describe this as the “shadow of the future,” where the possibility of future interactions makes cooperation more valuable and defection more costly (Dixit and Nalebuff 1991). As a result, even self-interested actors may find cooperation to be the better long-term strategy.

Game theorists describe several punishment strategies that help sustain this cooperation. One common strategy is tit-for-tat, where a player begins by cooperating and then mirrors the other player’s previous action in future rounds. If the other player defects, the first player retaliates by defecting in the next round but returns to cooperation if the other player cooperates again. Another strategy is grim trigger, where a player cooperates until the other defects once, after which they defect permanently (Dixit and Nalebuff 1991). Both strategies make defection costly over time because short-term gains from betrayal are outweighed by future losses from retaliation. Dixit and Nalebuff argue that these strategies work because repeated interactions allow players to punish defection in later rounds, making future cooperation more valuable than short-term betrayal (Dixit and Nalebuff 1991).

Dixit and Nalebuff also argue that repeated interaction allows players to build reputations, making future cooperative relationships more valuable than immediate gains from defection (Dixit and Nalebuff 1991). Wolford’s discussion of the “live-and-let-live” system in the trenches of World War I illustrates this theory in practice. British and German soldiers repeatedly encountered one another across the same trenches and gradually developed informal truces because violations could be punished in future encounters (Wolford 2019). Mengel’s meta-analysis of Prisoner’s Dilemma experiments provides additional empirical support, showing that cooperation is consistently higher in repeated games, particularly when participants are uncertain about when the interaction will end (Mengel 2018). Together, these theoretical, historical, and experimental perspectives demonstrate that repeated interaction can sustain cooperation by making future relationships more valuable than immediate gains from defection.

However, repeated interaction does not always lead to cooperation. As Wolford explains in his discussion of the arms race between Britain and Germany before World War I, both countries repeatedly increased military spending even though mutual restraint would have reduced costs for both sides (Wolford 2019). Despite the repeated interactions, fear of vulnerability and lack of trust led both countries to continue escalating. Unlike the trench truces, the cost of being exploited in the arms race was potentially catastrophic because falling behind militarily threatened national security. This made cooperation far riskier than defection. As a result, repeated interaction failed to sustain cooperation because the incentives

to avoid vulnerability outweighed the long-term benefits gained from cooperation. These examples show that repetition can encourage cooperation, but only when the structure of the interaction makes future cooperation more valuable than the immediate risks associated with defection.

In many settings, cooperation fails not because individuals are irrational, but because the incentives are structured in a way that makes defection the best individual choice. Consequently, even when mutual cooperation produces the best collective outcome, each player has a private incentive to defect. Defection either provides the highest possible payoff if the other player cooperates, or it protects against the worst outcome if the other player defects. This is the logic behind the dominant strategy in the Prisoner's Dilemma. Cooperation breaks down not because people do not understand the benefits, but because the incentive structure rewards defection (Shepsle and Bonchek 1997).

Furthermore, this problem becomes more severe when changing the subject from two-person interactions to larger groups. In group settings, people face a collective action problem, where everyone benefits from cooperation but each person is incentivized to take advantage of the contributions of others. Shepsle and Bonchek explain that as the number of participants increases, the incentive decreases. The contribution of each participant becomes less significant, and it becomes harder to monitor who is contributing and who isn't (Shepsle and Bonchek 1997). As a result, people may choose not to contribute, expecting others to do most of the work. This leads to an outcome where cooperation would not occur even though everyone would benefit from it.

Shepsle and Bonchek illustrate this problem through the tragedy of the commons. This occurs when a resource is shared among many individuals, where each person is incentivized to use as much of it as possible while the costs are distributed across the group. Over time, this leads to depletion of the resource, even when it would be beneficial for everyone if usage were limited. This example demonstrates how individually rational behavior can lead to collectively harmful outcomes when incentives are driven by self-interest (Shepsle and Bonchek 1997).

A different version of this problem occurs during wartime alliances. As Wolford explains in his analysis of World War I alliances, Austria had very little incentive to contribute fully to the joint war effort because Germany was going to fight regardless. Since Austria could benefit from Germany's efforts without contributing as much itself, it chose to focus on its own campaign rather than supporting the broader alliance strategy. This is an instance of free-riding in a collective action problem. Unlike the traditional two-player Prisoner's Dilemma, the issue here is not mutual defection but rather unequal contribution. However, the underlying problem is the same: incentives that lead to the breakdown of cooperation (Wolford 2019).

These examples show that cooperation often fails because the structure of incentives encourages defection or free-riding. Whether in small interactions or in large groups, individuals tend to respond to the incentive structures they face. When the incentives reward behavior that helps themselves, cooperation is difficult to sustain. Given these strong pressures toward defection, the key question becomes what factors can overcome these incentives and allow cooperation to persist. While cooperation is difficult to sustain, the rate at which it occurs varies across strategic settings. Experimental studies of

strategic interactions show that some people cooperate at high rates or at low rates, even when the situation or setup seems similar. This suggests that cooperation is not just a matter of whether individuals are willing to cooperate, but rather depends on the specific conditions of the interaction.

Mengel's analysis of experimental studies shows that cooperation changes based on several key factors. One important factor is the number of rounds in a game. Cooperation tends to be higher in repeated interactions, especially when participants do not know exactly when the interaction will end. When the last interaction is uncertain, people have stronger incentives to maintain cooperation because future interactions are possible. In contrast, when participants know the exact number of rounds, cooperation tends to decline toward the end, as individuals understand that there will be no future consequences for defection (Mengel 2018).

Another important factor is the payoff structure of the game. When the temptation to defect is high, meaning that individuals gain significantly from betraying a cooperating partner, cooperation rates tend to decrease. On the other hand, when the cost of being exploited is relatively low, individuals are more willing to cooperate. This shows that even small changes in incentives can produce large differences in behavior. The same basic game can lead to very different outcomes depending on how the payoffs are structured (Mengel 2018).

Information also plays a critical role in shaping cooperation. When individuals have access to information about their partners, such as past behavior or reputation, they are more likely to cooperate. Similarly, when individuals know how many rounds remain or have better information about the structure of the interaction, they can adjust their strategies accordingly. These factors influence expectations about others' behavior, which in turn affects decisions about whether to cooperate (Mengel 2018).

Dawes and Thaler provide more evidence that cooperation depends on context. In their review of public goods experiments, they explain that participants usually contribute about 40–60% of their resources in one-shot games (Dawes and Thaler 1988), even though standard game theory predicts that they should contribute nothing. However, cooperation changes when the game is repeated. In repeated public goods experiments, cooperation often starts relatively high but declines over time as participants observe others free-riding and adjust their own behavior. This shows that cooperation is not a fixed personal trait; it responds to the structure of the situation, including whether the game is one-shot or repeated, whether people expect others to contribute, and whether they feel taken advantage of (Dawes and Thaler 1988).

Looking beyond the Prisoner's Dilemma also highlights how cooperation varies across different types of strategic interactions. In collective action problems, such as wartime alliances, cooperation depends on incentives related to contribution and free-riding rather than mutual defection. In coordination games, cooperation is easier because players share a common goal but must align their actions. These different types of games demonstrate that cooperation is not a single behavior, but rather a set of behaviors shaped by the structure of the interaction and the context surrounding it. Factors such as repetition, payoff structure, information, and game type all influence whether cooperation emerges or collapses. However, structural conditions alone do not fully explain cooperative behavior. Psychological processes also shape

how individuals perceive strategic situations and respond to incentives, which introduces another layer of variation in cooperation.

The previous sections focused on structural and contextual explanations for cooperation, such as repetition, incentives, and structure of the game. These explanations all assume that individuals respond rationally to the setup of the game. However, psychology plays a major role in the behavior of cooperation. Even when individuals are in the same exact strategic situation, they may respond differently based on how they think about the decision and how quickly they need to respond. This suggests that cooperation is not just a rational response to incentives. In many cases, people may be psychologically inclined towards cooperation.

One important explanation comes from research on intuition and decision making. Rand and his coauthors conducted experiments in which participants played cooperative games under different time constraints. They found that participants were more likely to cooperate when they had less time to think about their decisions. When under pressure on time, cooperation rates increased, while giving participants more time to deliberate reduced cooperation (Rand et al. 2012). This finding suggests that, in these experimental settings, cooperation may often be an instinctive or automatic response rather than a carefully calculated strategy. The results also indicate that participants were more likely to defect after having additional time to deliberate, although this pattern may depend on the strategic setting and the individuals involved.

Kahneman's dual-process framework helps explain why Rand's results occur. Kahneman distinguishes between two systems of thinking. System 1 is fast, automatic, and intuitive, while System 2 is slow, deliberate, and analytical (Kahneman 2011). Rand's experiments match the results with this concept. In the condition with a short time constraint, participants relied more heavily on System 1 thinking and cooperated at higher rates. In the condition with a long time constraint, participants followed more System 2 reasoning and became more likely to defect (Rand et al. 2012). Together, these findings suggest that cooperation can sometimes be an intuitive response, while deliberation may increase the likelihood of defection in certain strategic settings. Rand provides the experimental evidence, while Kahneman provides the psychological explanation for why these patterns are visible.

These findings qualify one of the central assumptions of traditional game theory by suggesting that human behavior cannot always be explained entirely through strategic reasoning. Although Rand's experiments indicate that intuition can promote cooperation in some settings, the evidence does not imply that intuition always produces cooperation or that deliberation invariably leads to defection. Instead, intuitive psychological processes appear to shape decisions before individuals consciously evaluate incentives and payoffs.

Psychology influences cooperation not only through intuition but also through motivation. Andreoni introduces the concept of "warm-glow" altruism to explain why individuals cooperate even when there are no external incentives encouraging them to do so. Andreoni distinguishes between two forms of altruism. In pure altruism, individuals care about the overall group outcome. They want the public good to be provided, regardless of who contributes. If someone else contributes enough, a purely altruistic person

is satisfied because the collective outcome is achieved. In warm-glow altruism, however, individuals receive utility from the act of contributing itself. Cooperating produces a private psychological reward independent of the actual effect on the group outcome (Andreoni 1990).

This distinction is important because it introduces an internal incentive for cooperation. In a one-shot Prisoner's Dilemma, a purely strategic player should defect because there are no future rewards, punishments, or reputational benefits tied to cooperation. However, a player motivated by warm-glow altruism receives an additional psychological payoff simply from cooperating. The act of cooperating itself has value. This internal reward can make cooperation worthwhile even when the external incentive structure favors defection. External incentives aren't the only path to cooperation because the player carries an internal one (Andreoni 1990).

Andreoni's explanation differs from the mechanisms discussed by Rand and Kahneman. Rand and Kahneman focus on how individuals make decisions, specifically the difference between intuitive and deliberate thinking. Andreoni instead focuses on why individuals choose to cooperate in the first place. His argument is not about the thinking process but about motivation. Individuals may cooperate because doing so aligns with their personal values or provides emotional satisfaction.

DISCUSSION

One important connection across the literature is the relationship between structural and psychological explanations for cooperation. At first glance, these mechanisms appear separate. Sections 1–3 focused on structural factors such as repeated interaction, punishment, reputation, and incentives, while Section 4 focused on intuition and altruism. However, the evidence suggests that these mechanisms are actually interdependent. Rand's experiments show that individuals often cooperate instinctively in one-shot games, even when defection would maximize their payoff (Rand et al. 2012). Yet Dawes and Thaler demonstrate that cooperation frequently declines over repeated rounds as participants observe others free-riding (Dawes and Thaler 1988). If cooperation is truly intuitive, it should remain stable across interactions. Instead, the evidence suggests that intuition may help initiate cooperation, but structural mechanisms such as punishment and reputation are often necessary to sustain it. Cooperation is therefore strongest when psychological motivations and structural incentives reinforce one another.

A second tension emerges between the game-theoretic and psychological explanations for cooperation. Dixit and Nalebuff argue that cooperation can be sustained through deliberate punishment strategies such as tit-for-tat and grim trigger (Dixit and Nalebuff 1991). These strategies assume that individuals carefully calculate future consequences and adjust their behavior accordingly. However, Rand's findings suggest that deliberate thinking often reduces cooperation (Rand et al. 2012), while Kahneman's distinction between System 1 and System 2 thinking suggests that cooperation is more likely when decisions are made intuitively (Kahneman 2011). At first glance, this creates a contradiction within the literature. The game-theoretic explanation assumes that cooperation emerges through careful calculation, while the psychological explanation suggests that deliberation can reduce cooperation in certain settings. However,

these mechanisms may operate at different stages of an interaction rather than directly contradict one another. Intuition may encourage individuals to cooperate initially, while repeated interactions allow strategic calculation, punishment, and reputation to sustain or modify cooperation after individuals observe one another's behavior. Rather than competing explanations, these mechanisms may complement one another by explaining different phases of cooperative decision making.

The literature also demonstrates that context is not simply a background condition but a factor that determines which mechanisms can operate effectively. Wolford provides two examples involving repeated interaction that produce opposite outcomes. In the World War I trenches, repeated interaction allowed soldiers to develop a live-and-let-live system that reduced violence and sustained cooperation. In contrast, repeated interaction failed to prevent the arms race between Britain and Germany before World War I (Wolford 2019). Mengel's findings help explain this difference. Cooperation is influenced by contextual variables such as observability, attributability, and the ability to respond to defection (Mengel 2018). In the trenches, violations were immediately visible and retaliation was possible. In the arms race, military expansion was more difficult to monitor and far more difficult to reverse. These differences suggest that repetition alone does not determine whether cooperation succeeds. Instead, context determines whether the mechanisms that support cooperation can function at all.

These tensions lead to an important open question. The analysis above suggests that intuition and structure are interdependent, that psychological and game-theoretic explanations partially contradict one another, and that context determines which mechanisms can operate effectively. This raises a question that the paper cannot fully answer: in any given strategic interaction, how can researchers determine which mechanism is doing the most work? The comparison between the World War I trenches and the pre-war arms race suggests that context may often be decisive, but additional research comparing a wider variety of strategic settings would be necessary to determine whether this pattern holds more broadly.

CONCLUSION

This paper examined the question of what factors sustain cooperation despite the predictions of classical one-shot game-theoretic models. Section 1 showed that repeated interaction can sustain cooperation by making future punishment and reputation more valuable than short-term gains from defection. Section 2 demonstrated that incentive structures can either encourage or undermine cooperation, particularly in collective action problems where individuals benefit from free-riding. Section 3 revealed that cooperation varies significantly depending on contextual factors such as payoff structures, repetition, and expectations about others' behavior. Finally, Section 4 introduced a psychological dimension, showing that cooperation is often influenced by intuition and altruistic motivations rather than purely strategic calculation.

Taken together, these findings suggest that cooperation cannot be explained by any single mechanism. Instead, cooperation emerges through the interaction of structural mechanisms, contextual conditions, and psychological motivations. The predictive failures of traditional game theory often stem from treating

these factors in isolation rather than recognizing how they reinforce or weaken one another across different strategic settings.

This analysis has several limitations. Much of the experimental evidence examined in this paper comes from laboratory studies involving university students, raising questions about whether these findings generalize to real-world interactions with higher stakes and longer time horizons. Additionally, many of the historical examples discussed, including World War I trench warfare, the pre-war arms race, and Omertà, are drawn from Western and European contexts. Cooperation may operate differently in societies with different cultural norms and social structures. This paper also does not examine institutional enforcement as a mechanism for sustaining cooperation, despite the important role that laws, treaties, and formal organizations often play in encouraging cooperative behavior.

The tensions identified throughout this paper point toward an important question for future research. If cooperation can emerge through both structural incentives and psychological motivations, how can researchers determine which mechanism is doing the most work in a particular strategic setting? Comparing a wider variety of cases, including non-Western contexts and situations involving formal institutional enforcement, may provide a clearer understanding of the conditions under which cooperation emerges or breaks down.

REFERENCES

- Andreoni, James. "Impure Altruism and Donations to Public Goods: A Theory of Warm-Glow Giving." *The Economic Journal*, vol. 100, no. 401, 1990, pp. 464–77. <https://doi.org/10.2307/2234133>.
- Campana, Paolo, and Federico Varese. "Cooperation in Criminal Organizations: Kinship and Violence as Credible Commitments." *Rationality and Society*, vol. 25, no. 3, 2013, pp. 263–89. <https://doi.org/10.1177/1043463113481202>.
- Dawes, Robyn M., and Richard H. Thaler. "Anomalies: Cooperation." *Journal of Economic Perspectives*, vol. 2, no. 3, 1988, pp. 187–97. <https://doi.org/10.1257/jep.2.3.187>.
- Dixit, Avinash K., and Barry J. Nalebuff. *Thinking Strategically: The Competitive Edge in Business, Politics, and Everyday Life*. W. W. Norton, 1991.
- Kahneman, Daniel. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.
- Mengel, Friederike. "Risk and Temptation: A Meta-Study on Prisoner's Dilemma Games." *The Economic Journal*, vol. 128, no. 616, 2018, pp. 3182–3209. <https://doi.org/10.1111/econj.12548>.
- "Mob Commission Trial: Mob Boss Carmine Persico Conducts His Own Defense." *Archives & Special Collections*, University of Virginia Law Library,

<https://archives.law.virginia.edu/dengrove/writeup/mob-commission-trial-mob-boss-carmin-persico-conducts-his-own-defense>. Accessed 7 June 2026.

Rand, David G., Joshua D. Greene, and Martin A. Nowak. "Spontaneous Giving and Calculated Greed." *Nature*, vol. 489, 2012, pp. 427–30. <https://doi.org/10.1038/nature11467>.

Shepsle, Kenneth A., and Mark S. Bonchek. *Analyzing Politics: Rationality, Behavior and Institutions*. W. W. Norton, 1997.

Wolford, Scott. *The Politics of the First World War: A Course in Game Theory and International Security*. Cambridge University Press, 2019.