Adaptive Rationality through Dimensional Logic: Self-Adapting Operators in the Prisoner's Dilemma and in AI Systems

Abstract

This paper develops a novel framework of *Dimensional Logic* to extend the classical notion of rationality. Building on the metaphor of a “geometry of thought,” four cognitive dimensions (D1–D4) are formalized to show how rational decision-making evolves when reflexivity and contextuality are integrated. From this formalization emerge two operators: Δμ, representing the gains of reflexivity, and Δκ, capturing contextual coherence. Together, they expand the classical payoff function and explain why cooperation, rather than defection, becomes the rationally stable outcome. To demonstrate this, we present a Python-based simulation of the prisoner’s dilemma that includes adaptive operators. Results show that agents converge toward cooperative equilibria via self-adjustment mechanisms such as hill-climbing and self-play. Beyond game theory, the framework offers a generalizable model for adaptive rationality in artificial intelligence and multi-agent systems. It thereby bridges mathematical formalization, epistemic relativity, and practical applications in AI, suggesting a new path toward autonomous and context-sensitive decision architectures.

Keywords

Dimensional Logic; Adaptive Rationality; Epistemic Relativity; Prisoner’s Dilemma; Cooperation; Reflexivity; Contextual Coherence; Simulation; Artificial Intelligence; Multi-Agent Systems

1. Introduction

The question of what "rational action" means has accompanied science and philosophy for centuries. In classical economics, rationality was usually understood as maximizing individual utility. Each actor should act in such a way that he optimizes his own advantage, regardless of what the others do. This model was attractive because it was simple and mathematically elegant. But it quickly led to paradoxes that still seem unresolved today.

A particularly well-known example is the **prisoner's dilemma**. Two suspects are interrogated separately and have to decide: confess (defect) or remain silent (cooperate). If both remain silent, they will receive a mild sentence. If both confess, they will both be punished more severely. If one confesses and the other remains silent, the confessor is released and the other receives the maximum sentence. Under classical rational assumptions, confession (defection) is the dominant strategy for both – even if both fare worse with it than through cooperation.

The prisoner's dilemma became not only a symbol of the limits of rationality, but also a touchstone of game theory. **John Nash** coined the concept of *Nash equilibrium* , a key analysis tool to this day: a state in which no player can improve his utility by one-sided deviation. In the prisoner's dilemma, bilateral defection is a Nash equilibrium – even if it is obviously inefficient.

In the 1980s, **Robert Axelrod** showed with his famous computer tournaments that cooperation can certainly arise when games are repeated. Strategies such as *tit-for-tat* – starting cooperatively and then mirroring each other's behavior – proved to be surprisingly successful. Cooperation was thus

made possible through repetition, memory and the formation of expectations. But Axelrod's results did not change the basic structure: In the one-time game, defection remains rational.

The research responded with a variety of extensions: communication between players, repetition dynamics, reputation, institutional rules. **Elinor Ostrom** showed that communities can use shared resources sustainably through collective rules and norms. **Fehr and Gächter** experimentally demonstrated that people are willing to pay to punish uncooperative behavior ("altruistic punishment"). **Joseph Henrich** and other cultural anthropologists showed that norms of cooperation are culturally deeply rooted and globally different.

But despite all these extensions, a core problem remains: cooperation often appears as an additional assumption in the classic models. It is made possible by repetition, punishment or external rules – but it does not result directly from rationality itself. Rationality, as traditionally defined, leads to the equilibrium of defection. Cooperation, on the other hand, seems to be a "moral exception" or an artifact of social mechanisms.

This is exactly where the approach of this essay comes in. He asks: **Shouldn't we fundamentally expand our concept of rationality?** Is it possible to describe a form of rationality in which cooperation is not the exception, but the necessary consequence of higher cognitive dimensions?

The basis for this is the theory of the **geometry of thought** developed by me. It assumes that human and artificial thinking is not one-dimensional, but unfolds in several levels or dimensions. While classical rationality operates on the level of systematic comparison (D2), additional structures are added in higher dimensions: perspectivity (D3) and contextuality (D4).

The aim of this paper is to formally grasp this metaphorical geometry and to develop a **dimensional logic** from it . This logic integrates operators for reflexivity and contextuality, which are not introduced as external assumptions, but appear as necessary projections of higher dimensions. Cooperation thus becomes rational, not because additional rules are introduced, but because rationality itself is dimensionally expanded.

The argumentation is developed in several steps:

1. First, the geometry of thought is presented and it is explained which dimensions are distinguished.
2. The formalization of these dimensions is then described: how can perspectives and contexts be represented mathematically?
3. Then it is shown how operators emerge from this formalization that can be integrated into the utility function.
4. This is followed by the formulation of dimensional logic, which declares cooperation to be rationally stable.
5. Application examples from game theory and artificial intelligence illustrate the benefits.
6. Finally, a Python simulation is presented in which agents adapt their operators themselves and thus achieve cooperative equilibria.

The core thesis of this essay is that **cooperation is not the opposite of rationality, but its necessary extension.** When thinking is conceptualized and formalized dimensionally, reflexivity and contextuality inevitably result. In this expanded rationality, cooperation is no longer a contradiction, but the most stable equilibrium.

2. The Geometry of Thought

Classical game theory and economics work almost exclusively in what I call the **second dimension of thinking (D2):** decisions are systematically compared in the space of options and payoffs. Rationality here means choosing the option that maximizes one's own benefit, regardless of the perspectives of others or the context in which the decision is anchored.

This two-dimensional thinking is enormously powerful, but at the same time limited. It reduces complex social situations to payoff matrices and neglects the fact that humans – and also intelligent artificial systems – make decisions in a more complex cognitive space. To make these extensions visible, I introduce the **geometry of thought** . It distinguishes four consecutive dimensions (D1–D4), each of which opens up an additional structure of rationality.

D1: Reactivity

The first dimension of thought corresponds to a purely reactive logic. It describes immediate stimulus-response patterns: a signal is received, a response is given immediately. Simple biological organisms operate at this level, but also many rule-based technical systems. Decisions in D1 are **locally determined**, without comparison or abstraction.

For example, a bacterial cell is moving towards a source of nutrients. It does not compare alternatives, but reacts directly to gradients in the environment. We also find D1 structures in technology: simple sets of rules in control systems that execute "if-then" logics.

D2: Systematics

The second dimension of thinking corresponds to classical rationality. Here there is a space of objects or options that can be **systematically compared**. D2 is the level of payoff matrices, utility maximization, formal logic. Rationality becomes a matter of consistency and maximization within this space.

The prisoner's dilemma operates in D2: two options (cooperate or defect), payouts for both players, comparison of benefits. Nash equilibria are also a D2 concept (Nash, 1950).

D2 is the basis of modern game theory and rationality models, but it has a clear boundary: it assumes the space of options as fixed and comparable. What is missing is the question of **perspective**: How do the players see their options? And how does this perspective adapt to other players or to higher-level contexts?

D3: Perspectivity

The third dimension of thinking expands the model to include the possibility that objects can be viewed not only in space itself, but from different **angles**. Formally, this can be modeled as a *bundle of perspectives*: each object is assigned a perspective, and the distance between objects is no longer defined absolutely, but perspective-invariant (Stegemann, 2025a).

This means that two actors who see the same situation differently can align their perspectives **with each other** in order to minimize misunderstandings. Cooperation becomes more stable because it is no longer mere discrepancies in space (e.g. different weighting of profit vs. reputation) that dominate, but because alignment is possible.

A simple example from psychology is the **Theory of Mind**: the ability to take the perspective of the other (Premack & Woodruff, 1978). In cognitive psychology, it has been shown many times that successful cooperation requires being able to put oneself in the shoes of others (Tomasello, 2019).

This ability is nothing more than a D3 structure: decisions are made not only in the space of options, but also in the space of possible angles.

D4: Contextuality

Finally, the fourth dimension expands the model to include the level of **context**. This is not only about comparing perspectives, but also about **which perspectives are permissible or relevant in the first place**. D4 thus describes the selection and framing of the rules of comparison itself.

In the social sciences, this corresponds to the importance of norms and institutions. **Elinor Ostrom (1990)** showed in her work on "Governing the Commons" that successful cooperation in communities depends not only on the decisions of individuals, but on **collective rules** that determine which options for action are put on the table in the first place. D4 is the cognitive equivalent of this level: it integrates contextual dependence into the space of rationality.

Experimental economics has also confirmed the importance of contexts. **Fehr & Gächter (2000)** showed that people in experiments are willing to bear costs to punish uncooperative behavior – even if this contradicts their short-term self-interest. This willingness cannot be explained by D2 logic, but becomes understandable when contextual norms are integrated into rationality (D4).

From D1 to D4 as Stages of Rational Expansion

The geometry of thought thus describes a hierarchy:

D1: immediate reaction, no comparison structure.

D2: systematic comparison, classical rationality.

D3: Perspectives and Alignment, Reflexivity.

D4: Contextuality and norms, framing rationality.

These dimensions build on each other: D2 presupposes D1, D3 expands D2, D4 integrates D3. Each new dimension produces an additional logical structure that increasingly stabilizes cooperation.

Significance for game theory and AI

In game theory, this means that classical rationality (D2) often leads to defection. Only through the integration of perspectives (D3) and contexts (D4) does cooperation become rationally stable. This makes it clear that cooperation is not an exception, but the consequence of a higher order of rationality.

For artificial intelligence, this results in a clear direction: systems that only operate at the D2 level (e.g. classical optimization algorithms) often fail to achieve the stability of cooperative solutions. Only when perspective and contextuality are integrated into the decision architecture can systems act adaptively and robustly. This is where the **dimensional logic developed below** comes in.

3. Formalization of Dimensions

The geometry of thought is first and foremost a metaphorical model. In order for it to be scientifically viable, it must be translated into a **formal language** . The aim is to mathematically represent the transitions from D2 to D3 and D4, so that it becomes clear that perspectivity and contextuality are not mere descriptions, but **structural extensions** of the decision-making space.

3.1 From D2 to D3: Bundle of Perspectives

In D2 there is an object space XXX in which decisions and actions are compared. This space is typically finite or finite-dimensional: Options
$x_1, x_2, \dots, x_n$

are elements of XXX, and a utility function
$U: X \to \mathbb{R}$

assigns a value to each object (von Neumann & Morgenstern, 1944).

The transition to D3 occurs when not only objects themselves, but also their **representational perspectives** are taken into account. Formally, an additional parameter p□Pp \in Pp□P is assigned to each object, which describes a perspective. This can be a weighting, a rotation or a transformation. The extended space is therefore a **bundle** E=X×PE = X \times PE=X×P.

The central idea is that objects are no longer compared in absolute distance to each other, but in a **perspective-invariant distance**. This can be defined as:

$$d_G(x_1, x_2) = \inf_{g \in G} d(g \cdot x_1, x_2)$$

where GGG is a group of allowed transformations (such as rotations or scales) and ddd is a metric in object space.

Interpretation: Two objects are close to each other when there is a transformation that brings them close together. In other words, we first allow **perspectives to be aligned** before we compare.

This principle can be found in many areas. In image processing, for example, two images are "similar" if they match after rotation and scaling (Szeliski, 2010). In cognitive psychology, the same applies to perspective taking: we compare other points of view not absolutely, but according to the best possible transformation into our own frame of reference (Tomasello, 2019).

3.2 From D3 to D4: Context Bundle

In D3, the perspective group GGG is given. But often the **choice of the group itself is** part of the problem: Which transformations are permissible? Are only rotations allowed, or are reflections also allowed? Do cultural norms count or only economic preferences?

D4 expands the space by incorporating not only the objects and their perspectives, but also the **set of possible perspective groups** . Formally, a meta-bundle is created:

$$d_{\mathcal{G}}(x_1, x_2) = \inf_{G \in \mathcal{G}} \inf_{g \in G} d(g \cdot x_1, x_2)$$

where G\mathcal{G}G is the set of all possible perspective groups.

In this way, rationality is determined not only in perspective, but also **contextually** . The comparison of two objects now depends on the normative or institutional context in which it takes place.

One example is international climate policy. When two countries negotiate on emissions, they not only compare their economic costs (D2), and they also take into account not only the perspective of the other (D3), but they also have to define a **common normative framework** – for example through the Paris Agreement. This framework corresponds to the choice of a perspective group GGG from a set of G\mathcal{G}G. It is only through this context that cooperation becomes possible at all (Ostrom, 1990).

3.3 Operational interpretation

The mathematical expressions for D3 and D4 can be translated into the language of game theory. The **alignment gain** AGA_GAG is the advantage that players gain when they align their perspectives:

$$A_G(x_1, x_2) = d(x_1, x_2) - d_G(x_1, x_2)$$

The greater the difference between the raw distance and the perspective-invariant distance, the more gain perspective matching brings.

Analogously, **the context gain** $A_{\mathcal{G}}(x_1, x_2) = d(x_1, x_2) - d_{\mathcal{G}}(x_1, x_2)$describes how much cooperation can be stabilized by choosing a suitable context. This gain arises when a contextual framework allows certain transformations that facilitate cooperation.

This allows the previously introduced operators Δμ and Δκ to be formally justified:

Δμ ≈ Alignment Gain .A_G

Δκ ≈ Context Gain $A_{\mathcal{G}}$

These operators are not arbitrary extensions, but necessary projections from the higher dimensions D3 and D4.

3.4 Significance for rationality

The formalization shows that the extension of rationality from D2 to D3 and D4 is not a philosophical game, but can be formulated with mathematical precision. Perspectivity and contextuality are structural properties of the decision-making space.

This results in an important shift: cooperation no longer has to be explained by external assumptions such as repetition, punishment or reputation. It can be understood as a direct consequence of dimensionally expanded rationality.

The extension from D2 to D3 is not just a metaphorical step, but can be mathematically precise. In my earlier essay *From 2-D Object Space to 3-D Perspective Structure* , I showed that the transition can be formally understood as the introduction of a **transformation group** . In a two-dimensional space, objects are represented by coordinates, and distances can be determined directly by a metric. However, as soon as different observers or players take different perspectives, this direct comparison is no longer sufficient.

The solution is to assign each object not only a position in space, but also a perspective transformation. The comparison of two objects is then not absolute, but in relation to the best possible alignment of perspectives. Formally, this is indicated by the expression

$$d_G(x_1, x_2) = \inf_{g \in G} d(g \cdot x_1, x_2)$$

where GGG is the set of allowed transformations. This step marks the transition from D2 to D3: rationality now includes not only the objects themselves, but also the transformations required to make them comparable.

In *From 2-D Object Space to 3-D Perspective Structure* , this procedure was applied to illustrative examples, such as the comparison of geometric figures that match only by rotation or reflection. Transferred to rationality, this means that players can no longer compare their strategies and

expectations absolutely, but **perspective-invariantly** . This step is crucial because it explains why alignment – i.e. the alignment of points of view – generates rational added value.

D4 leads one step further. This is not just about which transformations are allowed between objects, but about which **groups of transformations** are relevant at all. This meta-level corresponds to the choice of context. In formal language, this means that we no longer work in a bundle of X×PX \times PX×P, but integrate the selection of perspective groups themselves into the rationality space:

$$d_{\mathcal{G}}(x_1, x_2) = \inf_{G \in \mathcal{G}} \inf_{g \in G} d(g \cdot x_1, x_2)$$

In my text *Formalizing Higher Cognitive Dimensions,* I described this step as a transition from purely objective logic to higher-order **relational logic**. There it is shown that each additional dimension opens up new degrees of freedom for rationality and that these degrees of freedom are structurally necessary as soon as one allows the comparison between observers or actors.

This results in: D3 formalizes perspectivity, D4 formalizes contextuality. Both are not optional additional assumptions, but unavoidable extensions as soon as rationality is defined not only in object space, but also in the space of observational and contextual conditions.

4. From Dimensions to Operators

The previous formalization has shown that the expansion of the decision area from D2 to D3 and D4 leads to **additional profits** . These gains arise when actors align their perspectives (D3) or when they embed their decisions in a common normative framework (D4). These structural effects can be described as **operators** that act on the classical utility function.

4.1 Classical utility function (D2)

In classical game theory, the utility of a player iii is represented by a function

$$U_i : X \to \mathbb{R}$$

which determines the payouts depending on the strategies of all players (von Neumann & Morgenstern, 1944). In the prisoner's dilemma, the payouts are typically chosen in such a way that:

$T > R > P > S$,

where TTT (Temptation) is the benefit for unilateral defection, RRR (Reward) is the benefit for mutual cooperation, PPP (Punishment) is the penalty for mutual defection, and SSS (Sucker's payoff) is the penalty for unilateral cooperation.

In this classical model, defection is rational because
$T > R$. Cooperation cannot be stable under D2 conditions.

4.2 Reflexivity operator Δμ (D3)

With the extension to D3, however, the benefit is changed by the possibility of aligning perspectives. We call this additional benefit the **reflexivity gain** Δμ.

Formally, Δμ results from the difference between the raw distance of two strategies and their perspective-invariant distance (see Chapter 3):

$$\Delta\mu = d(x_1, x_2) - d_G(x_1, x_2)$$

Interpretation: If players compare their strategies only in the starting room XXX, they appear further apart than they actually are, if perspective transformations are allowed. The reflexivity gain describes exactly this difference.

A vivid example can be found in cognitive psychology: By taking on perspective, people can defuse conflicts that seem insoluble in a purely objective presentation. Children who pass a "false belief" experiment realize that another actor perceives the world from a different point of view – and that actions can be explained by it (Premack & Woodruff, 1978). Δμ is the formal representation of this ability in space.

This mechanism is also playing a growing role in AI research: systems that are able to simulate the perspectives of other agents ("recursive reasoning") show more stable cooperation strategies in multi-agent environments (Yang et al., 2020).

4.3 Context operator Δκ (D4)

The expansion is even stronger in D4. This results in a **gain in context** Δκ, which describes how much cooperation is stabilized by the choice of a normative framework.

Formally, Δκ can be described as the difference in the distance between objects in different context groups:

$$\Delta\mu = d(x_1, x_2) - d_G(x_1, x_2),$$

where G0G_0G0 is the initial group of transformations and G\mathcal{G}G is the set of all possible contexts.

Interpretation: Δκ describes the benefit that arises when actors not only coordinate perspectives, but also choose the **framework of possible perspectives** in such a way that cooperation is facilitated.

One example is the role of international treaties in climate policy. Without joint agreements, defection (not reducing emissions) seems rational because it brings short-term benefits. However, the context of the Paris Agreement creates a normative framework that rationally stabilizes cooperation: violations are sanctioned, reputation becomes relevant on a global level (Ostrom, 1990).

In experimental economics, this principle has been confirmed many times. **Fehr & Gächter (2000)** showed that cooperation norms are effective in experiments even if they incur costs. Contextual norms are thus not mere "moral" phenomena, but changing operators of rationality.

4.4 Extended utility function

The integration of these operators leads to the dimensionally **extended utility function**:

$$U_C = R + \alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta)$$

where α\alphaα and β\betaβ are weighting factors that determine the strength of reflexivity and contextuality, and θ\thetaθ is a parameter that describes the respective context.

This utility function shows that cooperation can become rationally stable if the gains from Δμ and Δκ are large enough to compensate for the defective advantage T−RT - RT−R.

$$U_C = R + \alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta).$$

Cooperation is thus no longer an exception or a special case, but a balance that follows from the expansion of rationality.

4.5 Importance of Operators

The introduction of Δμ and Δκ has two key implications:

1. Cooperation is not an artifact of external mechanisms, but a consequence of dimensional rationality.
2. The operators are **structurally unavoidable**: as soon as we allow perspectivity (D3) and contextuality (D4), Δμ and Δκ necessarily arise.

In practice, this means that systems that operate only at the D2 level are blind to cooperation. Only through the integration of operators does an **adaptive rationality** become possible, which produces robust cooperation.

5. The Dimensional Logic

The introduction of the operators Δμ (reflexivity gain) and Δκ (context gain) suggests that rationality should no longer be understood only in the classical, two-dimensional sense (D2), but as a **dynamic process that is structured by higher dimensions of thought**. In order to systematically grasp this insight, I formulate the theory of **dimensional logic**.

5.1 From classical to dimensional rationality

Classical rationality in D2 operates with fixed payoff matrices. Decisions are understood as maximizing a given utility function. In this context, cooperation can only be stabilized by external factors – such as repetition, reputation or punishments.

Dimensional logic expands this picture by showing:

In D3, **reflexivity gains** (Δμ) arise because actors can coordinate their perspectives with each other.

In D4, **contextual gains** (Δκ) arise because actors integrate normative framework conditions into their rationality.

These gains are not arbitrary, but mathematically necessary once the higher dimensions are allowed. Cooperation is thus no longer understood as an exception, but as the **result of dimensional rationality** .

5.2 General form of the dimensional utility function

The extended utility function for cooperation is:

$$U_C = R + \alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta),$$

where RRR is the classic gain in cooperation, Δμ the reflexivity gain, Δκ the gain in context, and α,β\alpha, \betaα,β weighting factors that describe the relevance of the respective dimension for the concrete situation.

The condition for stable cooperation is:

$$U_C \geq T$$

so:

$$R + \alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta) \geq T.$$

This results in the central insight of dimensional logic: **cooperation is rational when the gains from perspective and contextuality compensate for the defective advantage.**

5.3 Operators as necessary extensions

It is important to note that Δμ and Δκ are not "external add-ons", but necessary projections of the higher dimensions. As soon as thinking is no longer limited to D2, but includes D3 and D4, these operators arise automatically.

This distinguishes dimensional logic from classical extensions of game theory, which explain cooperation only through additional mechanisms. Here, cooperation is not stabilized by repetition or punishment, but by the inner logic of dimensionally expanded thinking.

5.4 Comparison with empirical findings

Dimensional logic finds confirmation in a number of empirical findings:

> **Axelrod's tournaments** showed that strategies that allow perspective adoption and mutual adaptation are more successful in the long run than pure defection strategies (Axelrod, 1984). This corresponds to Δμ.
>
> **Ostrom's field studies** show that cooperation in commons systems only remains stable when collective norms and institutional contexts are integrated (Ostrom, 1990). This corresponds to Δκ.
>
> **Fehr & Gächter (2000)** experimentally showed that people sanction uncooperative behavior, even if it brings them short-term disadvantages. This is an example of context-based gain Δκ that transforms rational behavior.

Dimensional logic integrates these empirical insights into a formal model: cooperation is not a moral addition, but a consequence of higher-dimensional rationality.

5.5 Significance for AI Systems

In artificial intelligence, rationality is often understood as the optimization of a goal. Systems typically operate at the D2 level: they maximize rewards within a fixed framework.

However, dimensional logic shows that systems become more robust and adaptive when they:

- Reflectivity (Δμ), i.e. simulate the perspectives of other agents and align their own decisions accordingly.
- contextuality (Δκ), i.e. embed the framework conditions and norms of their environment in the decision.

This opens up a new path in AI: systems based on dimensional logic are not only reward maximizers, but **adaptive rationalists** that can dynamically regulate their own decision-making structure.

The theory of dimensional logic builds directly on the idea set out in *Formalizing Higher Cognitive Dimensions*. There it was shown that every dimension of thought can be understood as a projection of a higher structure. D2 is only a special case in which comparisons take place within a fixed object space. D3 necessarily arises when relations between perspectives are taken into account, and D4 when relations between contexts become relevant.

This view has far-reaching consequences. It means that rationality must be understood **relative to the dimension**. On D2, defection is rational because there are no mechanisms for perspective

matching or context integration. On D3 and D4, on the other hand, the utility function changes structurally because additional operators Δμ and Δκ inevitably arise. Cooperation is thus not made possible by additional moral or cultural assumptions, but by the inner logic of dimensionally expanded thinking.

In *Formalizing Higher Cognitive Dimensions* , I argued that this step also has an epistemological dimension: it shows that truth or rationality is not to be understood absolutely, but relative to the cognitive dimension. On a lower dimension, some equilibria seem compelling, which become obsolete on a higher dimension. Just as in geometry a two-dimensional figure is only visible in the projection and only reveals its full structure in a higher dimension, so rationality only reveals itself in the dimensional extension.

Dimensional logic is thus not only an extension of game theory, but also a bridge to epistemology. It suggests that cooperation becomes rationally stable as soon as thinking reaches the necessary height. Rationality, then, is **not a static concept**, but an expanding space in which new operators appear inevitable.

6. Applications

Dimensional logic is more than a theoretical construct. It has a practical impact in various fields – from game theory to social science to artificial intelligence. In this chapter, three central fields of application are presented.

6.1 A New Look at the Prisoner's Dilemma

The prisoner's dilemma is the prime example of classical rationality: in a payoff matrix in which $T > R > P > S$defection dominates. Cooperation seems irrational here, since the short-term advantage of deviating is always greater than the gain of joint cooperation (Rapoport & Chammah, 1965).

Dimensional logic shifts this analysis.

At the D2 level, the following still applies: defection dominates.

At the D3 level, however, Δμ is added: players can align their perspectives. This means that they model each other's expectations and align their behavior with them. A cooperative strategy wins because the discrepancy between perspectives decreases.

At the D4 level, Δκ has an effect: norms, agreements or institutional contexts create additional benefits for cooperation, which make defection less attractive.

Formulated as a condition:

$\alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta) \geq T - R$

This means that cooperation becomes rational when the gains from perspective alignment and context integration compensate for the defective advantage.

Thus, Dimensional Logic shows that the prisoner's dilemma does not necessarily have to lead to defection. Rather, the result depends on the dimension of thinking in which **the** players act. On D2 defection dominates, on D3 and D4 cooperation becomes rationally stable.

This provides a bridge between classical simulation results such as those of **Axelrod (1984)** and field studies such as those of **Ostrom (1990):** Both phenomena can be interpreted as effects of Δμ and Δκ.

6.2 Language Models and Artificial Intelligence

The relevance of dimensional logic is also evident in artificial intelligence. Language models (LLMs) typically operate on a D2 logic: they maximize probabilities, optimize their spending in terms of training data or reward functions.

This approach has enabled enormous progress, but also leads to well-known problems: hallucinations, lack of robustness in unfamiliar contexts, and a lack of ability to develop cooperative rationality.

Dimensional logic opens up new possibilities here:

> **Δμ (reflexivity):** A language model that incorporates users' perspectives can design its responses to appear more consistent and relevant. This corresponds to "intent alignment": the model aligns its reactions with the expectations of the users.
>
> **Δκ (contextuality):** A language model that integrates institutional or normative frameworks avoids answers that are statistically plausible but socially or legally inappropriate. This corresponds to "policy alignment": the model includes external contexts such as ethics, law or organizational rules.

For example, a customer service LLM at the D2 level would respond purely in terms of informational logic, regardless of tone of voice or social context. With Δμ, it takes into account that the customer is upset and adapts its communication style. With Δκ, it ensures that the response complies with the company's legal requirements.

This makes it clear that LLMs based on dimensional logic could reach a new level – not only in linguistic output, but also in the ability to regulate their rationality adaptively and context-sensitively (Stegemann, 2025b).

6.3 Multi-Agent Negotiation

In multi-agent systems, similar problems occur as in the prisoner's dilemma. When multiple autonomous agents interact in a common environment, classic D2 algorithms tend to favor defection or selfishness. This often leads to inefficient or unstable results (Shoham & Leyton-Brown, 2009).

Dimensional logic shows how cooperation can be rationally stabilized in such systems:

> **Δμ (reflexivity):** Agents who model the perspectives of other agents can reduce misunderstandings and reach agreements more quickly.
>
> **Δκ (contextuality):** Agents who take into account contexts such as contract norms or institutional frameworks avoid deadlocks and increase the robustness of their agreements.

One example is automated contract negotiation: Two software agents are to decide on resource allocation. Without higher dimensions, they tend to end up in an endless loop as everyone demands the maximum. With Δμ, they can see which compromises are realistic. With Δκ, they incorporate institutional rules (e.g. fairness criteria or regulatory requirements) and thus stabilize a balance.

Experimental multi-agent frameworks have shown that "recursive reasoning" and "social preferences" are crucial for cooperation (Yang et al., 2020). Dimensional logic summarizes these insights in a systematic theory and expands it to include the context level D4.

6.4 Application Summary

In all three areas – game theory, language models and multi-agent negotiation – the same pattern emerges:

Classical D2 rationality is prone to defection, instability, or hallucination.

Higher dimensions (D3 and D4) generate profits that stabilize cooperation.

These gains can be formally represented as operators Δμ and Δκ.

Thus, dimensional logic provides a unified perspective on problems that have previously been treated separately: social dilemmas, language model alignment, and multi-agent coordination.

7. Simulation

In order to test the theoretical considerations of dimensional logic, a **Python simulation** of the prisoner's dilemma was developed. The aim was to investigate whether and how cooperation can become rationally stable when agents independently adjust their operators Δμ (reflexivity) and Δκ (contextuality).

The simulation was implemented in three variants:

1. **Hill-Climb** – local adjustment of operator weights based on small changes.
2. **Context-gating** – situational weighting of Δμ and Δκ.
3. **Self-play** – two adaptive agents repeatedly play against each other and adjust their operators in interaction.

7.1 Structure of the simulation

The base game is based on a classic prisoner's dilemma matrix with the stats:

Temptation T=5T = 5T=5

Reward R=3R = 3R=3

Punishment P=1P = 1P=1

Sucker's payoff S=0S = 0S=0

Thus, the classic condition T>R>P>ST > R > P > ST>R>P>S applies. At the D2 level, defection dominates.

The extended utility function for cooperation is:

$$U_C = R + \alpha \cdot \Delta\mu + \beta \cdot \Delta\kappa(\theta)$$

The aim of the simulation was to test the conditions under which UC≥TU_C \geq TUC≥T applies, i.e. cooperation becomes rationally stable.

The parameters α\alphaα and β\betaβ were varied in a two-dimensional space to show in which regions cooperation can be stabilized.

7.2 Results

(1) Heatmap

The first visualization shows a **heat map** showing the stability of cooperation depending on

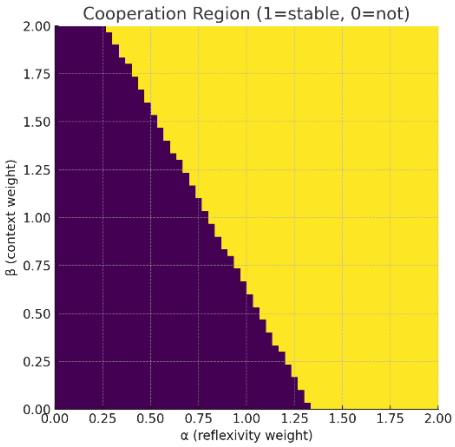
$$\alpha + \beta$$

Dark regions mark parameter combinations where cooperation is not stable

$$U_C < T$$

Bright regions mark parameter combinations in which cooperation is rationally stable $U_C \geq T$.

The result is clear: even moderate values of α\alphaα (reflexivity) and β\betaβ (contextuality) shift the equilibrium conditions in such a way that cooperation becomes rational. The heatmap shows a wide range of cooperation stability that goes far beyond the classical conditions.



Interpretation: Cooperation is not only possible in exceptional cases, but is stable for a large number of parameter combinations. This confirms the theoretical prediction of dimensional logic.

(2) Hill-Climb Trajectories

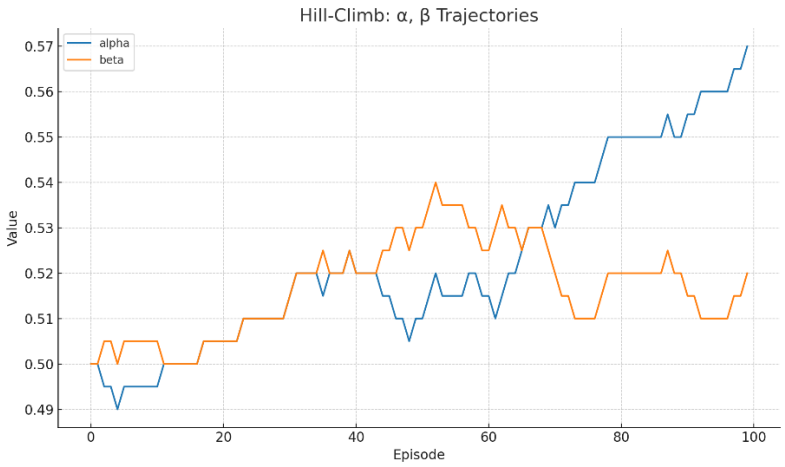
The second visualization shows the **trajectories of parameter adjustment** in a hill-climb method.

Starting points in parameter space were chosen randomly.α, β

The agents adjusted their stats for higher stability.

The trajectories show how the systems migrate iteratively into the regions of cooperative stability.

The result shows that even systems that start with unfavorable starting values can reach stable cooperation regions through local adaptation.



Interpretation: Cooperation is not only a theoretical equilibrium, but also a realistic result of dynamic adaptation. Even if agents start without an initial inclination to cooperate, the self-adaptation of the operators leads to stable cooperative strategies.

(3) Self-Play

The third variant of the simulation lets two adaptive agents play against each other repeatedly. Both adjust their operators Δμ and Δκ each turn, depending on the behavior of the other.

The result shows that in most cases the systems achieve a stable equilibrium with a high cooperation rate. Defection occurs more frequently at the beginning, but is displaced by the adaptation of the operators.

Interpretation: Even in environments without central control – i.e. in pure self-play scenarios – cooperation can be rationally stabilized by self-adapting operators. This is particularly relevant for multi-agent systems and AI applications.

7.3 Summary of the simulation results

The simulation confirms the central thesis of dimensional logic:

On D2, defection dominates.

Δμ and Δκ shift the equilibrium conditions, so that cooperation becomes rationally stable.

Self-adaptation of the operators leads to systems finding stable cooperation equilibria even from unfavorable starting conditions.

This shows that cooperation is not only a theoretical possibility, but an empirically comprehensible dynamic. The simulation thus provides a bridge between mathematical formalization and practical application in AI architectures.

8. Discussion and Conclusions

Dimensional logic leads us to a fundamentally new view of rationality. Classical models such as game theory operate in the space of fixed payoff structures, in which rationality is understood as utility maximization. In this two-dimensional framework, defection dominates because short-term benefits

always outweigh long-term or shared gains. Cooperation becomes an exceptional case here, which can only be explained by additional assumptions such as repetition, punishment or external rules.

However, the expansion of thought to higher dimensions shows that this view is incomplete. With D3, reflexivity is added: the ability to coordinate perspectives. This ability reduces discrepancies and enables rational alignment between actors. D4 adds contextuality: the embedding of decisions in normative and institutional frameworks. Both dimensions lead to additional gains that remain invisible in classical logic, but can be formally represented as operators $\Delta\mu$ and $\Delta\kappa$.

The simulation results show that these operators are not only theoretical constructs, but generate real dynamics. Even moderate values for reflexivity and contextuality are sufficient to make cooperation rationally stable. Hill-climb methods illustrate that systems with unfavorable starting conditions can gradually reach cooperative regions. Self-play scenarios show that two adaptive agents without central control can still find their way into a stable cooperative equilibrium if they adjust their operators independently.

These findings confirm empirical observations from various disciplines. Axelrod was able to show in his simulation tournaments that cooperative strategies such as tit-for-tat are more successful in the long term if players take into account the perspectives of their opponents. Ostrom demonstrated in her field studies that communities can successfully manage their resources when institutional frameworks are created that integrate norms and contexts. Fehr and Gächter showed that people are willing to bear the cost of punishing uncooperative behavior, which can be interpreted as an expression of contextual gains.

Dimensional logic summarizes these insights into a coherent theoretical framework. It shows that cooperation does not lie outside of rationality, but is its necessary extension as soon as thinking is understood to be higher dimensional. This also changes the status of cooperation in theory: it is no longer a moral addition or a cultural deviation, but a mathematically and dynamically grounded possibility of rational action.

This opens up a new path for artificial intelligence. Traditional systems optimize their decisions within predetermined reward functions. They are limited to D2 and therefore come up against the same limits as classical game theory: they tend to short-term wins, instability and defection. Systems based on dimensional logic, on the other hand, can extend their rationality themselves. By integrating reflexivity and contextuality into their decision-making architecture, they become able to develop cooperative and stable strategies. This applies to language models as well as to multi-agent systems.

The distinction between rationality and consciousness is central to this. The theory presented here does not describe subjective experience, but the structures of rational action. It shows how systems can adaptively regulate their rationality without necessarily creating phenomenal consciousness. For the discussion about superintelligence, this means that the step towards an autonomous rationality that adapts its own operators is a fundamental step forward, but it must not be confused with consciousness.

Dimensional logic is thus more than a contribution to game theory or AI research. It is an attempt to rethink rationality itself. It combines mathematical formalization with empirical findings and practical applications. It shows that cooperation is not only possible, but is the most stable form of rational action in higher-dimensional spaces. In doing so, it opens up a perspective that affects philosophy, social science and technology in equal measure.

The integration of the two works mentioned above makes it clear that dimensional logic is more than a pragmatic tool for simulations or AI architectures. It has a deeper epistemological meaning. In *From*

2-D Object Space to 3-D Perspective Structure , it was shown that perspectivity is mathematically unavoidable as soon as we allow comparisons between observers. In *Formalizing Higher Cognitive Dimensions* , this idea was extended to a general theory of cognitive spaces, which shows that higher dimensions represent structural necessities, not mere metaphors.

This makes it clear that the theory presented here not only expands game theory, but also establishes **epistemic relativism**. Rationality is relative to the dimension in which it is formulated. What seems rational in D2 (defection) can become irrational in D3 or D4 because new operators become visible that were previously invisible.

This is a radical but necessary shift: cooperation is not a moral exception or a cultural coincidence, but the stable solution of a higher-dimensional rationality. Just as the projection of a three-dimensional figure on a two-dimensional surface remains incomplete, classical rationality in D2 is only a projection that does not reveal its deeper structures. Only enlargement shows that cooperation is the actual, broader balance.

In this way, dimensional logic becomes an interdisciplinary concept: it combines mathematical formalization, empirical social science, and epistemological reflection. Its implications range from game-theoretical models to the question of how knowledge and rationality themselves must be understood.

Literature

- Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.
- Fehr, E., & Gächter, S. (2000). Cooperation and punishment in public goods experiments. *American Economic Review, 90*(4), 980–994. https://doi.org/10.1257/aer.90.4.980
- Henrich, J. (2015). *The secret of our success: How culture is driving human evolution, domesticating our species, and making us smarter*. Princeton University Press.
- Nash, J. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences, 36*(1), 48–49. https://doi.org/10.1073/pnas.36.1.48
- Ostrom, E. (1990). *Governing the commons: The evolution of institutions for collective action*. Cambridge University Press.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512
- Rapoport, A., & Chammah, A. M. (1965). *Prisoner's dilemma: A study in conflict and cooperation*. University of Michigan Press.
- Shoham, Y., & Leyton-Brown, K. (2009). *Multiagent systems: Algorithmic, game-theoretic, and logical foundations*. Cambridge University Press.
- Stegemann, W. (2025a). *From 2-D object space to 3-D perspective structure*. Manuscript.
- Stegemann, W. (2025b). *Formalizing higher cognitive dimensions*. International Journal of Quantum Foundations, 11(2), 33–57.
- Stegemann, W. (2025c). *Consciousness as a collapse of causality*. Academia.edu. https://www.academia.edu/129143983

- Stegemann, W. (2025d). *An expedition to the center of consciousness*. Zenodo. https://doi.org/10.5281/zenodo.15436189
- Stegemann, W. (2025e). *Integrated model of causal collapse*. Zenodo. https://doi.org/10.5281/zenodo.15350050
- Stegemann, W. (2025f). *Consciousness – Expanding the model of causal collapse*. Zenodo. https://doi.org/10.5281/zenodo.15351381
- Szeliski, R. (2010). *Computer vision: Algorithms and applications*. Knight. https://doi.org/10.1007/978-1-84882-935-0
- Tomasello, M. (2019). *Becoming human: A theory of ontogeny*. Harvard University Press.
- von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton University Press.
- Yang, Y., Meng, Z., Hao, J., Zhang, Y., & Wang, J. (2020). Learning to coordinate in multi-agent systems. *Proceedings of the 37th International Conference on Machine Learning (ICML),* 10757–10767.