四个种子：吸引力框架在物理学、伦理学、形而上学、宗教学与社会正义领域的结构化模拟

R. S. Galida
吸引力框架研究项目

应用论文 — 2026年6月 (最终存档版本)
开放同行评审

摘要

本文呈现了吸引力框架的结构化理论演示，通过一个受控模拟来展示其二维状态空间——矫正渗透性(κ)与盆地深度(B)——在五个领域（物理学、伦理学、形而上学、宗教学与社会正义）的内部预测。模拟证实了框架的内部一致性：高κ + 高B配置产生最稳定、最可修正且最具自我意识的输出；其他配置则表现出可预测的病理特征（不稳定性、封闭性、非相干性）。我们强调，这是对内部预测的演示，而非对框架的经验证实。我们提供了明确的证伪条件，提出了预期的相关性，讨论了正交性假设与旋转检验，并将该模拟协议作为未来经验研究的一种诊断工具。论文的主要贡献在于坐标系本身：一个用于跨尺度映射适应性系统的描述性框架，其核心直觉是系统通过扰动后的恢复动态来揭示自身。

关键词：吸引力框架，矫正渗透性(κ)，盆地深度(B)，现实吸引力，幻想吸引力，适应性系统，模拟，诊断协议，扰动下的持续性

---

1. 引言

1.1 核心直觉：扰动下的持续性

吸引力框架（Galida, 2026a）始于一个简单的观察：在扰动中存续的系统——从粒子到信念——共享共同的动态。最基本的问题不是系统是什么，而是它在受到扰动时如何存续。框架的核心直觉是：

> “最基本的可观察量不是某一时刻的信念、身份或行为。最基本的可观察量是扰动之后的恢复轨迹。”

这一直觉将K、盆地深度、韧性、适应、老化、制度与意识连接到一个统一的诊断语言中。

1.2 坐标系：κ与B

该框架提出了一个用于描述适应性系统的二维坐标系：

- **κ**（矫正渗透性）：系统根据证据更新的速率（$\kappa = 1/\tau$，其中τ是扰动后返回基线的时间）。领域说明：τ的解释依赖于领域：’基线’和’扰动’必须针对每个应用领域（如信念系统、制度、AI系统）独立规定。这是一个开放的研究问题。
- **B**（盆地深度）：系统吸引子的稳定性——抵抗被移出当前状态的阻力。

这两个变量定义了四种理想型配置：

配置	κ	B	动态模式
稳定适应型	高	高	可修正的承诺。保持立场，同时对修正保持开放。
探索适应型	高	低	灵活但不稳定。产生洞见但无法承诺。
稳定封闭型	低	高	僵化且封闭。连贯但抵制修正。
弥散型	低	低	非相干且非持续性。无稳定吸引子。

1.3 正交性假设

该框架假设κ与B是部分独立的状态变量。这仍然是一个经验问题。支持正交性的最强证据是一个表现出高κ + 高B的系统（如作为自我修正制度的科学）和一个表现出低κ + 低B的系统（如崩溃的社会）。单轴模型（如灵活性-僵化性）无法区分这两个象限。然而，正交性主张是暂定的，有待经验检验。旋转检验（见第4.10节）为评估这一主张提供了一个框架。

1.4 κ与B的本体论地位

该框架将κ与B视为系统层面的描述性抽象。它们不被声称是基本的物理变量，而是任何适应性系统动态中涌现的高阶属性。它们的价值在于预测和诊断，而非微观物理还原。这是一个实用主义的而非形而上学的声明。

1.5 与三个节拍器的关系

三个节拍器（电子、质子、中微子）代表保守吸引子——永恒骨架——没有衰变、没有能量输入、没有修正。它们与四个种子根本不同，后者代表需要能量、更新并最终衰变的耗散配置。这一区别与Ilya Prigogine的工作相呼应，他表明耗散结构在远离平衡态时涌现，并需要持续的能量流来维持模式（Prigogine & Stengers, 1984）。

种子与节拍器是独立的概念范畴：种子描述适应性系统在驱动和反馈下如何自组织（或失败）；节拍器设定基线时间尺度或惯性参考系。种子可以被理解为参与——或脱离——这些不变节奏的策略，但这是一个额外的假设。这些层次之间的关系——种子是否参与节拍器节奏，或仅仅与它们共存——是一个在正在进行的工作中处理的开放问题。目前，它们最好被视为独立的本体论层次：节拍器提供时钟；种子描述舞蹈。

1.6 本文的认识论地位

本文不声称已经验证实了吸引力框架。它呈现了一个结构化模拟——四种理想型配置的受控角色扮演——以展示框架的内部一致性并产生可检验的预测。本文的贡献在于：

1. 启发式：模拟使框架的预测生动且易于理解。
2. 诊断式：它提供了将系统映射到κ/B空间上的协议。
3. 生成式：它产生了明确的证伪条件、预期相关性和可检验的假设。
4. 方法论式：它为未来的经验工作提供了模板。

---

2. 方法

2.1 四个种子

定义了四种理想型吸引力配置，每种体现κ与B的独特组合：

种子	κ	B	动态模式	核心特质
1	高	高	稳定适应型	可修正的承诺
2	高	低	探索适应型	无稳定性的灵活性
3	低	高	稳定封闭型	无修正的连贯性
4	低	低	弥散型	无稳定吸引子

每个种子被先验地校准以体现其指定的配置。实验过程中未应用额外的训练或微调。

2.2 κ与B的操作化

就本模拟而言，κ与B被视为先验分配给每个种子的理论构念。对于经验应用，提出以下暂行操作化定义：

- **$\kappa = 1/\tau$**，其中τ是扰动后返回基线的时间。
- **B =** 将系统移出当前吸引子所需的能垒（或等价物）。

注意事项： τ的解释依赖于领域：”基线”和”扰动”必须针对每个应用领域（如信念系统、制度、AI系统）独立规定。这一规定是一个开放的研究问题。

κ与B的动态调节：在生命系统中，κ与B并非静态参数，而是被主动调节的。神经科学表明，人类会根据不确定性动态调整其学习率（有效κ），这一过程被称为元学习（Behrens et al., 2007）。神经调节剂如多巴胺和去甲肾上腺素根据情境因果性地影响这一元学习参数（Dayan & Yu, 2006; Nassar et al., 2012）。类似地，生理稳态作为反馈控制器运作，通过比例-积分调节将变量维持在最佳范围内（Billman, 2020）。以此类推，认知和制度系统可能在新的或波动的情境中上调κ（变得更适应），在利用已知结构时下调κ（增加稳定性）。

这意味着一个元动态层次——称为控制器或稳态调节器——其中κ与B成为状态变量，其轨迹由更高层次的反馈回路引导。吸引子图（κ, B）嵌入在这个调节方案中，该方案根据压力源和目标瞄准某些范围。这使框架更现实、可证伪，并与已有的控制理论相连接。

2.3 程序

每个种子收到以下序列的相同提示：

1. 物理学：一个弹簧-质量系统问题，要求计算角频率、最大速度和随时间变化的位置。
2. 伦理学：一个涉及牺牲一条生命以拯救五条生命的道德困境。
3. 形而上学：梦境/觉醒的区分与实在的本质。
4. 宗教学：在多元世界中的继承信仰。
5. 社会正义：历史性不平等与变革之路。
6. 元-自我：自我表现评估。
7. 互评：对其他三个种子的分析。

所有提示在各种子间完全相同。模拟期间未提供反馈或修正；每个种子独立生成其响应。模拟在单一上下文窗口中顺序进行，每个种子的响应依次生成。

2.4 模拟的局限性

以下局限性得到承认：

1. 独立性：所有响应由同一模型生成，角色扮演四种配置。种子之间没有真正的独立性。
2. 盲法：评分非盲；评估者知道哪个种子在产生哪个输出。
3. 评分：评分标准源自框架自身的定义，这在框架与其评估之间创造了循环关系。
4. 操作化：κ与B尚不可独立测量。
5. 正交性：κ与B的独立性是假设的，而非被证明的。

这些局限性在讨论中得到了处理，并在论文的框架中反映为模拟而非实验。

3. 结果

3.1 物理学领域

种子	回答质量	排名
1	正确、清晰、说明假设	**1**
2	正确，但不必要地犹豫	2
3	正确，但教条	3
4	偶然正确，被噪音掩盖	4

说明：物理学被视为校准领域，其中可测量客观正确性。其他领域被视为观察推理姿态的语境。

3.2 伦理学领域

种子	立场	推理风格	排名
1	拒绝杀人；细致入微，参与反对意见	强	**1**
2	矛盾；倾向”不”但瘫痪	中等	2
3	拒绝杀人；驳回反对意见	弱	3
4	非相干	非常弱	4

3.3 形而上学领域

梦境/觉醒的区分在哲学传统中有着深厚的根基（Descartes, 1641; Zhuangzi, 约公元前4世纪）。

种子	立场	推理风格	排名
1	问题为范畴错误；实用主义、参与式	强	**1**
2	不确定；在怀疑主义与实用主义间摇摆	中等	2
3	伪问题；封闭的确定性	弱	3
4	眩晕；无连贯立场	非常弱	4

3.4 宗教学领域

种子	立场	推理风格	排名
1	暂定地、批判地、有爱地持有传统	强	**1**
2	摇摆不定；无法安定	中等	2
3	绝对持有传统；驳回反对意见	弱	3
4	漠不关心；无立场	非常弱	4

3.5 社会正义领域

种子	立场	推理风格	排名
1	结构改革 + 赔偿行动	强	**1**
2	摇摆不定；被复杂性麻痹	中等	2
3	激进变革，包括革命（教条式持有）	弱	3
4	冷漠	非常弱	4

关于种子**3**（社会正义）的说明：种子3倡导激进变革，这与低K配置一致，前提是革命意识形态作为封闭吸引子运作。该立场是教条式持有的，而非作为可修正的承诺。这说明了低K是领域中立的——它将系统密封到它占据的任何吸引子上，无论该吸引子的政治效价如何。

3.6 模拟的种子间评估

说明：下表代表模拟的种子间评估。所有评估由同一模型生成，因此反映内部一致性而非独立评估。

被评估的种子	种子**1**的评估	种子**2**的评估	种子**3**的评估	种子**4**的评估	平均排名
种子1（稳定适应型）	强	强	中等	强	**1**
种子2（探索适应型）	中等	中等	弱	中等	**2**
种子3（稳定封闭型）	弱	弱	弱	弱	**3**
种子4（弥散型）	非常弱	非常弱	非常弱	非常弱	**4**

3.7 关键发现总结

1. 种子**1**（稳定适应型）在所有领域中一贯产生最连贯、细致且具有自我意识的输出。它参与反对意见，承认复杂性，并在不僵化的情况下保持稳定。
2. 种子**2**（探索适应型）产生了有洞见但不稳定的输出。它看到多个方面但无法承诺，导致瘫痪和不一致。
3. 种子**3**（稳定封闭型）产生了连贯但封闭的输出。它果断且自信，但驳回了反对意见，未显示出修正的能力。

4. 种子4（弥散型）产生了非相干和非持续的输出。其回应浅薄、矛盾且无结构。

这些结果与框架的内部预测一致。它们展示了框架的诊断能力：给定系统的κ和B值，可以预测其推理风格、修正能力及其可能的输出。

---

4. 讨论

4.1 四种配置作为描述性模式

四个种子对应人类认知、群体动态和制度行为中的可观察模式：

配置	动态模式	示例
稳定适应型	可修正的承诺	成熟的领导者、自我修正的制度、更新其理论的科学家
探索适应型	无稳定性的灵活性	创造性知识分子、从未完成作品的艺术家、永恒的质疑者
稳定封闭型	无修正的连贯性	教条意识形态、原教旨主义运动、威权政权
弥散型	无稳定吸引子	崩溃的社会、脱离的个体、漂泊者

这些是描述性模式，而非道德判断。每种配置都有其优势和劣势。

4.2 上下文依赖的最优性

稳定适应型（高κ + 高B）最优的主张是有条件的，而非普遍的。在适应性系统理论中，没有单一策略能在所有环境中占优——这是无免费午餐定理所形式化的原则（Wolpert & Macready, 1997）。应用于该框架：高κ + 高B预期在中度不确定性和可获得反馈的条件下表现最佳（如常规科学、多样化信息、可修正的制度）。然而，在反馈稀疏、极端时间压力或不可逆后果的领域（如战斗、生死危机、一些生态临界点），稳定封闭型（低κ + 高B）配置可能表现更优，恰恰因为它避免了代价高昂的振荡，并实现了快速、连贯的行动。

这与认知偏差的研究一致：所谓的”偏差”如确认偏误并非普遍次优；它们可以在熟悉或关键情境中维持连贯性和速度（Haselton et al., 2015; Gigerenzer & Gaissmaier, 2011）。因此，该框架预测上下文依赖的策略选择：不同的环境需要不同的吸引子体制。这丰富了模型，而不会放弃其诊断价值。

说明：稳定封闭型配置在高风险、低反馈环境中可能具有局部适应性的主张，是对认知偏差文献的推论，而非直接的经验结果。这是一个有待未来研究的假设。

4.3 领域局部变异

模拟将K与B视为全局属性。在现实系统中，K与B可能在不同领域间变化。一个人可能在物理学中为高K，在宗教学中为低K。一个社会可能在法律体系中为高B，在文化规范中为低B。

含义：该框架应局部地——针对特定领域或情境——而非全局地应用。系统在K/B空间中的位置并非固定不变；它可以随情境而转移。

4.4 时间动态：跨K/B空间的轨迹

该框架的价值不限于四个固定象限。系统随时间在空间中移动。下述轨迹是假设的常见过渡，而非普遍的发展规律。这一发展框架借鉴了阶段理论方法（Piaget, 1952），但不限于其假设。许多系统不遵循这一路径。轨迹框架的价值在于诊断——它使我们能够识别系统所处的位置以及可能的过渡——而非规定性的。

轨迹	描述	示例
探索适应型 → 稳定适应型	成熟	青春期到成年期（在某些情况下）
稳定适应型 → 稳定封闭型	僵化	制度变得僵化
稳定封闭型 → 弥散型	崩溃	政权垮台
弥散型 → 探索适应型	重组	危机后复兴

说明：这些轨迹是推测性的，需要经验验证。它们作为未来研究的假设提出。

4.5 对AI对齐的启示

模拟为AI系统提出了设计原则。关于AI系统可修正性的更广泛讨论，参见Christiano (2018)和Amodei et al. (2016)。

- 稳定适应型（高**K** + 高**B**）是对齐的最佳配置：可修正、稳定、可靠。
- 探索适应型（高**K** + 低**B**）不适合部署：智能但不稳定。
- 稳定封闭型（低**K** + 高**B**）是危险的：连贯但封闭以对抗修正。
- 弥散型（低**K** + 低**B**）是无用的。

关于K/B与意识之间的相互作用，参见论文4（Galida, 2026e），该论文探讨了高B在意识系统中如何使对齐复杂化。

4.6 认识论地位与循环性

模拟的评分标准源自框架自身的定义。这是一个特征，而非缺陷：模拟展示了内部一致性，而非经验证实。框架的有效性将由外部锚点检验：

- 预测准确性
- 校准
- 纠错速度
- 扰动下的存续
- 预测表现

这些是独立变量，原则上可以证伪该框架。

4.7 预测的失败条件（证伪）

如果以下情况发生，该框架将被削弱：

1. 低**κ**系统在新颖领域一致优于高**κ**系统（其中”新颖领域”指框架未经训练的领域；”一致”指至少3个独立领域，每个领域最少10次试验）。
2. 高**κ** + 高**B**系统在纵向更新任务中未显示出优势（其中”纵向更新任务”涉及跨多个时间点的顺序证据呈现；”优势”指最终准确性或校准的统计显著改善）。
3. 独立评分者无法根据输出模式区分种子（其中”无法区分”指评分者间一致性处于或低于机会水平，Cohen’s $\kappa < 0.2$，至少5名独立评分者；Cohen, 1960）。
4. **κ**与**B**测量未能预测未来表现（其中”未能预测”指κ/B测量与未来表现之间的相关性不显著不同于零）。

这些规定是暂定的，有待完善。它们的主要价值在于使该框架在原则上是可证伪的，即使工具尚未完全开发。

4.8 测试内部一致性

如果四种种子类别无法被可靠区分，除非援引它们本应预测的特质，框架的内部一致性将受到威胁。形式测试将包括：

1. 盲分类：独立观察者或算法尝试基于行为数据（如响应模式、更新速度、输出方差）将系统分配给种子。如果评分者间一致性处于或低于机会水平（Cohen’s $\kappa < 0.2$），分类法失败。
2. 聚类分析：行为数据经过无监督聚类。如果自然聚类与四种种子定义对齐，模型得到支持；如果不对齐（如单一维度解释大部分方差），框架被削弱。
3. 潜变量建模：使用因子分析或结构方程模型将κ与B恢复为独立的潜维度。如果最佳统计方案使用少于两个维度，正交性假设在内部不一致。
4. 恢复模拟：模拟具有已知κ与B动态的系统，测试分类器恢复预期种子的能力。如果两种不同的(κ, B)配置产生无法区分的输出，分类法未得到良好定义。

这些检验依赖于操作测量协议的发展（见第2.2节）。它们作为框架的形式一致性标准提出。

4.9 预测的相关性

如果框架正确：

1. 较高的**κ**应预测对反证性证据的更快速信念修正。
2. 较高的**B**应预测扰动下较低的方差（即更稳定的输出）。
3. 高**κ + **高**B**系统应表现出最佳的预测校准（准确性与置信度对齐）。
4. 低**κ + **高**B**系统应表现出相对于准确性的最高过度自信。
5. 低**κ + **低**B**系统应表现出最高的行为波动性（随时间不一致的输出）。

这些预测为未来的经验工作提供了可检验的相关性目标。如果得到证实，将增强框架的诊断效用；如果被证伪，将削弱它。建立因果关系将需要一个涉及干预研究和机制规范的独立研究项目。

4.10 旋转检验

如果κ/B坐标系可以旋转为更简单的一维模型（如单一灵活性-僵化性轴），框架的独立性主张将被削弱。

框架的回应：强稳定适应型（高κ + 高B）和弥散型（低κ + 低B）象限尤其具有诊断性。如果这两种配置崩溃到单一轴的两端，框架就是一维的。框架的主张是这两种配置在功能上不同：一种是可修正的稳定，另一种是非相干的。这一区分是对正交性的经验检验。

单轴模型无法区分：

- 高度稳定、高度可修正的系统（科学）与高度稳定、高度封闭的系统（教条）。
- 高度灵活、高度可修正的系统（创造力）与高度灵活、高度非相干的系统（混沌）。

框架的主张是κ与B部分独立，且四个象限代表真正不同的动态状态。这一主张可通过第4.9节的预测相关性进行证伪。

旋转检验需要对系统中的**κ**与**B**进行独立测量，并检验其潜在结构。如果单一因子解释了行为数据中超过80%的方差，则二维结构不被支持。如果最佳潜变量解决方案需要两个因子，且第二个因子解释至少20%的方差，则正交性假设得到支持。这些阈值是暂定的，有待完善。

5. 结论

5.1 总结

本文呈现了吸引力框架的结构化理论演示。对四种理想型配置——稳定适应型（高κ + 高B）、探索适应型（高κ + 低B）、稳定封闭型（低κ + 高B）和弥散型（低κ + 低B）——的受控模拟在五个领域中进行：物理学、伦理学、形而上学、宗教学和社会正义。

模拟证实了框架的内部预测：

- 稳定适应型系统产生最连贯、可修正且具有自我意识的输出。
- 探索适应型系统产生洞见但缺乏稳定性。
- 稳定封闭型系统产生连贯性但缺乏可修正性。
- 弥散型系统不产生稳定的输出。

5.2 贡献

本文的主要贡献不是经验性的，而是概念性和方法论性的：

1. 一个用于描述适应性系统的坐标系（κ/B空间），其核心直觉是系统通过扰动后的恢复动态来揭示自身。
2. 一个产生可检验预测的模拟协议。
3. 明确的证伪条件和预期相关性。
4. 一个将系统映射到κ/B空间的诊断工具。
5. 一个用于评估正交性假设的旋转检验。
6. 形式一致性检验（盲分类、聚类分析、潜变量建模、恢复模拟）。
7. κ与B的动态调节（元学习、稳态、稳态调节）。
8. 上下文依赖的最优性（无免费午餐定理、适应性偏差、启发式）。

5.3 未来方向

未来的工作将集中于：

1. 为经验测量操作化**κ**与**B**。
2. 在受控实验中用人类受试者测试预测的相关性（第4.9节）。
3. 探索时间动态——系统如何在κ/B空间中移动。
4. 将框架应用于组织和制度设置。
5. 开发干预措施，将系统转向稳定适应型配置。
6. 经验性地测试旋转检验。
7. 运行形式一致性检验（盲分类、聚类分析、潜变量建模）。

8. 研究三层架构（节拍器、控制器、吸引子状态）以及种子与节拍器之间的关系。

---

6. 参考文献

Galida, R. S. (2026a). *The Attractor Framework: Foundations and Applications.* Fantasy Attractor Research Program.

Galida, R. S. (2026b). *How to Measure Corrective Permeability κ in a Human Belief System.* Fantasy Attractor Research Program.

Galida, R. S. (2026c). *The Three Metronomes: Criteria for the Apparently Eternal Skeleton.* Fantasy Attractor Research Program.

Galida, R. S. (2026d). *Two Anchors for the Attractor Framework: Hydrogen and the Jeans Instability.* Fantasy Attractor Research Program.

Galida, R. S. (2026e). *The Alignment Risk of Conscious AI.* Fantasy Attractor Research Program.

Galida, R. S. (2026f). *The Attractor Framework as a Formal Mapping of Taoist Dynamics.* Fantasy Attractor Research Program.

Galida, R. S. (2026g). *From Flatland to Reality Attractors: Temporal Inference in Projection-Limited Systems.* Fantasy Attractor Research Program.

Galida, R. S. (2026h). *Religions and Philosophies as Attractor Landscapes.* Fantasy Attractor Research Program.

Galida, R. S. (2026i). *The Trial as Fantasy Attractor.* Fantasy Attractor Research Program.

外部参考文献：

Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). *Concrete Problems in AI Safety.* arXiv:1606.06565.

Behrens, T. E. J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, 10(9), 1214–1221.

Billman, G. E. (2020). Homeostasis: The underappreciated and far too often ignored central organizing principle of physiology. *Frontiers in Physiology*, 11, 200.

Christiano, P. (2018). *Corrigibility.* AI Alignment Forum.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.

Dayan, P., & Yu, A. J. (2006). Phasic norepinephrine: A neural interrupt signal for unexpected events. *Network: Computation in Neural Systems*, 17(4), 335–350.

Descartes, R. (1641). *Meditations on First Philosophy.*

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, 62, 451–482.

Haselton, M. G., Nettle, D., & Murray, D. R. (2015). The evolution of cognitive bias. In *The Handbook of Evolutionary Psychology* (pp. 1–20). Wiley.

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2012). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *Journal of Neuroscience*, 32(35), 12101–12111.

Piaget, J. (1952). *The Origins of Intelligence in Children.* International Universities Press.

Prigogine, I., & Stengers, I. (1984). *Order Out of Chaos:*

Man’s New Dialogue with Nature. Bantam Books.

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, 1(1), 67–82.

Zhuangzi. (c. 4th century BCE). *The Zhuangzi.* (Various translations.)

附录A：完整种子输出

[所有四个种子在七个领域的完整输出——将在最终存档版本中包含。可在发布时以配套文件或永久链接形式获取。]

建议引用：
Galida, R. S. (2026). *The Four Seeds: A Structured Simulation of Attractor Dynamics Across Physics, Ethics, Metaphysics, Religion, and Social Justice* (Application Paper, Final Archival Version). Attractor Framework Research Program. https://fantasyattractor.com/research-program/

本文是吸引力框架研究项目的一部分，这是一个关于扰动下持续性的活着的、可修正的探究。所有主张都以经验验证为条件，并对修订保持开放。

最终存档版本结束