

收得回的那一遍

论一次练习能修改什么，不由它有多难、多重、多像正式场合决定，而由动作发起之后还能不能被收回决定

舞蹈研究 × 运动技能科学 × 学习心理学

作者 王德生 + Claude

栏目 学科通融 · 站外碰撞

篇幅

发表 2026年8月1日

摘 要

舞蹈排练里有一种练法叫「标记」：舞者以大幅削减的动作走一遍编舞，不跳、不转、不用力。行内的理由是省体力。可是实验测出来的结果是——**标记过的舞者，随后正式表演的质量高于全力练过的**。这个结果同时顶翻三家。运动技能领域说练习越接近正式情境越好，削减等于抽掉学习依据。学习心理学说容易的练习制造能力错觉，长期保持更差。舞蹈自己给的认知负荷解释，说不清为什么别的减负手段替代不了它。本文主张三家共用了一个从未被检验的预设——把练习当成同一动作按某个保真度打折的版本，于是只剩下像不像、难不难、重不重可调。撤掉这个预设，剩下的变量不是量，是**动作发起之后还能不能被收回**。本文把它命名为**撤回余量**，并给出一条可裁决的推论：**在不可撤回的区间里练结构，结构缺陷不会以结构缺陷的样子出现，它会以执行者能力不足的样子出现，并被随后的加练固化**。本文称之为执行遮蔽，并说明它不需要任何怪罪动机就会发生。文中交出两条使命题收窄的约束、六条各不相同的失败条件（其中一条只需一间排练厅与一份问卷）、以及一条写死日期的赌注。

这一篇由哪三个领域的争论撞成

三边对「一次练习被削减之后，学习会变好还是变差」给出互相排斥的答案：运动技能领域说练习必须保留正式情境的信息结构，削减即减损；学习心理学说变容易的练习会抬高当下表现而压低长期保持；舞蹈研究实测到削减之后反而更好，并把它归给认知负荷下降。三家的答案不能同真，而它们共用了**同一个未被检验的预设——练习是同一动作按某个保真度打折的版本**。本文另有一节交代四个领域早已看见同一形状却归错原因，来源见文末参考文献。

舞蹈研究 · 削减动作反而练得更好

[Warburton, Wilson, Lynch & Cuykendall, The Cognitive Benefits of Movement Reduction: Evidence From Dance Marking \(Psychological Science 24\(9\), 2013, 1732–1739\)](#)

舞者学习两段芭蕾组合，一段以正式速度全力练，另一段以标记练，随后两段都以正式方式表演并评分。结果是标记过的那一段表现更好。作者提出具身认知负荷假说：削减身体控制的负担，可以把资源释放给记忆与整合。按这条结论，做得更少反而学得更好。

<https://pubmed.ncbi.nlm.nih.gov/23863756/>

运动技能 · 练习必须代表正式情境

[Pinder, Davids, Renshaw & Araújo, Representative Learning Design and Functionality of Research and Practice in Sport \(Journal of Sport & Exercise Psychology 33\(1\), 2011, 146–155; 机构公开全文\)](#)

把「代表性设计」从实验方法学推广到练习设计，主张练习任务必须在信息结构上代表正式任务，而不只是在动作外形上相似；一旦把关键信息源从环境中抽走，动作的时序与控制会系统性改变。按这条结论，抽掉重量、惯性与真实位移的练习应当无效甚至有害。

https://www.aut.ac.nz/_data/assets/pdf_file/0007/380698/Representative-Learning-Design-and-Functionality-of-Research-and-Practice-in-Sport.pdf

[学习心理学 · 容易的练习制造能力错觉](#)

[Soderstrom & Bjork, Learning Versus Performance: An Integrative Review \(Perspectives on Psychological Science 10\(2\), 2015, 176–199\)](#)

系统梳理学习与表现的可分离性：练习当下的表现是长期学习的不可靠指标，某些安排会同时对两者产生相反效应；让当下更顺畅的做法常常削弱长期保持与迁移。按这条结论，标记这种把练习变得最容易的做法，长期保持应当更差。

<https://pubmed.ncbi.nlm.nih.gov/25910388/>

目 录

一、导论：最省力的那种练法，练出来的却更好.....	5
二、三种解释，以及它们为什么不能同时为真.....	6
三、三家共享的那个未被审查的预设.....	6
四、被归零的那一组维度做了什么.....	7
五、归零的选择不是任意的.....	8
六、为什么它决定这一趟能改哪一层.....	9
七、二维辨别表.....	10
八、空转：看起来在练，其实那一层没有出现.....	11
九、执行遮蔽.....	11
十、四个领域早已看见同一个形状，而它们各自归错了原因.....	13

十一、与最近的几种说法的分界	14
十二、逐领域兑现	16
十三、两条让步	16
十四、怎么测，与那个只需一间排练厅的检验	18
十五、反噬预言与适用边界	18
十六、本判断对它自己适用吗	19
十七、可以让本文失败的六件事，与一条写死日期的赌注	20
结 论	21
注 释	22

一、导论：最省力的那种练法，练出来的却更好

排练厅里最常见的一幕是这样的。一段刚编好的群舞走第三遍时，编导说：这一遍标一下就行。于是全场的动作忽然缩小。大跳变成脚下点两下，双圈转变成手指画一个圈，托举变成两个人碰一下手掌。音乐照常走，队形照常换，只是身体不再离地。

行内把这叫作标记。它从来不需要辩护，理由现成得几乎不像理由：省体力。一天六小时排练，不可能每一遍都跳满，把体力留到该用的时候，是常识。

2013 年，一组研究者做了一件在行内看来近乎多余的事：他们把这个权宜之计当成一个可以被检验的干预来测。舞者学习两段芭蕾组合，一段以正式速度全力练，另一段以标记练，随后两段都以正式方式表演并被评分。结果是标记过的那一段更好。

这个结果值得停一下。它不是「标记也不错」，也不是「体力受限时标记是合理的折中」。它是：**在同等练习次数下，做得更少的那一组，随后做得更好。**

如果只在舞蹈内部看，这个发现容易被消化成一句行话——标记帮助记路。但把它放到三个学科的当代主张之间，它就不再是一个可以被消化的发现，而是一个三面受敌的反常。

第一，运动技能领域近二十年最有影响的练习设计原则是代表性学习设计。它主张练习环境必须保留正式环境中的信息与动作之间的耦合关系。一旦把关键信息源从环境里抽走，动作的时序与控制就会系统性改变。用发球机代替真人发球，练出来的东西迁移不回去。按这条原则，标记抽走的东西多得惊人：重量、惯性、落地冲击、真实的空间位移、与同伴的真实距离。它应当是练习设计的反面教材。

第二，学习心理学这一侧的主流是合意困难，以及「表现不等于学习」这条区分。它主张练习当下的表现是学习的不可靠指标。许多让当下更顺畅的安排会削弱长期保持，许多让当下更别扭的安排反而增强它。标记是把练习变容易的最彻底的一种做法。按这条原则，它至多改善当下的顺畅感，长期保持应当更差。而实验测的恰恰是随后的正式表现，且更好。

第三，舞蹈研究自己给出的解释是认知负荷：身体不必承担复杂的控制负担，腾出的资源可用于记忆与整合。这个解释顺理成章，却有一处对不上。如果起作用的是负荷总量，那么用别的方式把负荷降下来应当同样有效。而排练厅里的常规做法恰好提供了对照——打拍子、报数、把长句拆成四小节分开练、先在脑子里过一遍。这些都在降负荷，也都在用，但没有哪一种被认为可以替代标记。行内的说法很直白：拆开练完还得连起来标一遍。

三家不能同时对。第一家预测标记无效甚至有害。第二家预测它短期好、长期差。第三家预测任何等效的减负荷手段都应当同样好。实测的是：有效，随后更好，且不可被其他减负荷手段替代。

本文从这个三面矛盾出发，但走的不是仲裁的路。仲裁需要有一方错。下面要说明的是：三家在同一个地方都没有错，也都不可能对，因为它们共用了同一个未经检验的预设，而反常正落在这个预设之外。

二、三种解释，以及它们为什么不能同时为真

先把三家摆到位，尽量按它们自己的最强形态摆。把对手削弱之后再反驳，得到的结论对谁都没有用。

代表性学习设计来自生态动力学的传统。它的原始关切是实验效度：在实验室里测出来的东西凭什么迁移到球场上。这一支把「代表性设计」从实验方法学推广到练习设计，主张练习任务必须在信息结构上代表正式任务，而不只是在动作外形上像。这是一个有实测支撑的强主张：用抛球机练击球，击球者的时序参数与面对真人投手时系统不同。它的立足点是——技能是人与环境的耦合，不是储存在人身上的一段程序；抽掉环境，剩下的那段东西根本不是要学的东西。

它到不了的地方在于：这条原则没有给「要学的是哪一层」留出位置。信息与动作的耦合是一个整体概念，它不区分「这一句该在第三拍起」与「这一跳需要多大的蹬地力」。两者在它的语言里都是耦合的一部分。于是它无法解释为什么抽掉一部分耦合，反而让另一部分学得更好。在它的框架里，抽掉就是抽掉。

合意困难来自记忆研究。它的立足点极为坚实：学习与表现在概念上与实测上都可分离，而人——包括教师、教练与学习者本人——系统性地用后者推断前者。这条洞见解释了大量教育实践中的顽固错误，也把「感觉练得顺」这条最常用的自评指标拆掉了。

它到不了的地方在于：难度在这个框架里是一个标量。合意与不合意的分界靠「是否引发更主动的编码加工」来划，而这条界线在实测中很难先验判断，通常只能事后按结果反推。挑战点那一支把它形式化为难度与学习者能力的匹配，仍然是一条单轴上的最优点。标记的问题在于：按任何一种把难度当量的读法，它都在轴的极低端，而它的结果在高端。要么它是那条曲线上的异常点，要么它根本不在这条轴上。

认知负荷离现象最近，也因此最容易被现象反噬。它主张标记降低了身体控制的负担，从而释放资源用于更高层的整合。这个说法有一个隐含的替代性承诺：负荷是可以从不同来源加减的同一种东西。而排练厅里的实际做法不支持这个承诺——同样降负荷的手段并不互相替代。

更麻烦的是，标记并非一律更轻。熟练舞者的标记里包含大量额外的、正式表演中根本不存在的动作：用手指代替整个身体的旋转，用一条手臂的路径代表另一个人的走位，边标边出声念自己的进出点。这些是新增的编码工作，不是被省掉的。认知民族志一侧把标记区分成「标给自己看」「标给别人看」「标给配合用」三种用途，这个区分本身就说明标记不是一种减法。

到这里可以做一个小结。三家分别站在环境耦合、难度调节、资源分配三个不同的位置上，各自都有实证支撑，各自都不能容纳这个反常。这种局面通常意味着：不是某一家的读数错了，而是三家在测同一个东西，而那个东西不是起作用的那个。

三、三家共享的那个未被审查的预设

把三家的语言并排放着看，会看到一个它们都在用、却没有一家把它当作可检验命题提出来的假定：

一次练习，是那个正式动作按某个保真度打折的版本。

代表性学习设计谈的是保真度中的「信息保真」，合意困难谈的是保真度调整带来的难度变化，认知负荷谈的是保真度下降释放出的资源。三种语言的差别在于关心保真度的哪一个侧面，共同点在于都把练习安置在同一条从「完全不像」到「完全一样」的连续轴上。正式演出是这条轴的右端点，纯想象是左端点，标记落在中间偏左。

这条轴上只能提三种问题：应该站在轴的哪个位置，这个位置对应的难度合不合适，这个位置省下的资源去了哪里。三种问题都合法，也都无法处理这样一个事实——**标记不是全力执行的一个缩小版。**

这一点值得说细。若标记是缩小版，那么它在每一个维度上都应当按比例衰减。实际不是。在标记里，下面这些维度维持在与正式演出完全一致的水平：动作的先后次序、每个动作的起止时刻、与音乐的对齐点、在场地上的路径与朝向、与其他舞者的相对位置与交汇时刻、每个短句的呼吸断句。而下面这些维度被压到接近零：力度、幅度、离地高度、与地面的冲击、身体承受的风险。

也就是说，标记不是「整体调到四成」，而是**把一组维度保留在满值，把另一组维度调到接近零**。这是一个选择，不是一个刻度。三家共用的那条轴上没有位置放「选择」这件事，因为在那条轴上，保真度是一个数。

这个观察本身还不构成一条新判断，它只是把问题重新提对了：**被归零的那一组维度，究竟做了什么？**若答案只是「省力气」或「减负担」，我们仍然停在量的语言里。下一节主张答案不在量里。

四、被归零的那一组维度做了什么

把力度、幅度、离地高度、冲击这一组维度归零，会产生一个副产品。这个副产品在既有的练习理论里从未被当作变量：**动作在发起之后仍然可以被收回。**

一个跳跃一旦蹬离地面，剩下的轨迹由离地瞬间的速度与角度决定，中途改不了。一次带着全身重量的转体一旦启动，惯性会把它带完。一个双人托举一旦把同伴举离地面，「不做了」这个选项在物理上已经不存在。而在标记里，这些动作都可以在任何一拍上停住：脚下点两下随时可以停在第二下之前，手指画的圈随时可以画一半停下，碰一下手掌随时可以不碰。

这个性质在实验心理学里有一个成熟的名字和一把成熟的尺子。**不可撤回点**指的是一个已经准备好的动作不再能被叫停的时刻，操作定义是成功抑制概率降到一半的那个时点。测它的范式有两支：停止信号任务，与预期反应抑制任务。

这里有一处值得指出的巧合。预期反应抑制任务的原始论文 1960 年发表在《Research Quarterly》，那是当时美国健康、体育与休闲协会的学刊。也就是说，**测量「动作还能不能被收回」的第一把尺子，出生在体育学科内部**，随后被认知心理学接管并精细化。而在它出生的那个学科里，它几乎从未被用于练习设计，只被用于研究反应抑制本身。

这把尺子今天已经能在近乎真实的运动情境中使用。有研究把预期反应抑制范式做成了篮球版本：受试者观看跳投视频，在球出手瞬间抬指封盖，其中四分之一的试次视频提前中止或呈现一个假动作，受试者应当收回动作。他们测得的不可撤回点落在球出手前约 187 毫秒，假动作条件下约 177 毫秒。同一组人随后把范式扩展到整体起跳的封盖动作，把「用整个身体去做」这个条件带进了实验室。他们引以为据的现实观察很说明问题：即便最高水平的防守者，在真实比赛里也有约七成的时候会被假动作骗起。

有了这把尺子，可以给出一个不依赖直觉的定义：

撤回余量：从动作发起到不可撤回点之间的时长，占该动作总时长的比例。取值在零与一之间。

它接近一，意味着这个动作几乎全程都还能被收回；接近零，意味着它一旦发起就只能走完。全力完成的大跳、真实的托举、全速的冲刺，它都极低；标记的同一段舞句，它极高。

这个量与三家关心的那三个量都不重合，这一点需要点明，否则它只是一个新名字。

它不是难度。一段极慢、极轻、极容易的动作可以有很低的撤回余量——把一杯满水递出去的最后一寸，动作既慢又轻，但松手之后收不回来。反过来，一段很难的动作也可以有很高的撤回余量——一段复杂的手部编排，做错了随时能停。

它不是与正式演出的相似度。一次全速的对抗性演练在相似度上极高，撤回余量极低。一次同样在场地上、同样有对手、同样有战术、只是约定不许身体接触的演练，相似度略降，撤回余量大幅上升。

它不是负荷。给节拍器、报数、把长句拆开，这些都降低负荷，但都不改变已发起的动作能否被收回。这正好解释了那个替代性难题：为什么其他减负荷手段不能替代标记。

五、归零的选择不是任意的

上一节说标记是「把一组维度保留在满值、另一组压到接近零」，这个说法还欠一步：为什么恰好是那一组被压到零？

看被压掉的那些——力度、幅度、离地高度、地面冲击、承受的风险。它们有一个共同性质：**它们都是把动作从可收回状态推向不可收回状态的原因**。力度决定了系统被注入多少动量，动量决定了惯性，惯性决定了「不做了」这个选项在什么时刻消失。离地是最极端的一种——脚一旦离开地面，与地面之间的力矩通道就断了，剩下的时间里身体对自己的轨迹几乎没有发言权。托举则把这个性质叠加到两个人身上：一方的不可撤回，把另一方的安全一并抵押了进去。

再看被保留的那些——次序、起止时刻、与音乐的对齐、路径与朝向、与他人的相对位置。它们的共同性质是：**它们全都是关系，而关系不需要力就能被摆出来**。用手指画一个圈就能表明「这里是一个向右的双圈转」，因为向右、双圈、在这一拍上，这三件事都是关系，不是量。

于是标记这件事可以被重新描述一遍，而且描述得比「削减」准确得多：**它把动作里所有会导致不可撤回的成分抽掉，同时把所有关系性的成分原样留下。**这不是一次减法，是一次分离。

行内那句「标一下就行」之所以听起来像省事，是因为被抽掉的那一组恰好也是最耗体力的那一组。省力是这次分离的副产品，不是它的目的，更不是它起作用的原因。这个巧合很可能就是它长期被记在「省体力」名下的原因。

这个说明还带来一对可检验的条件，而且它们把「省力」与「可撤回」彻底分开。

一边是**同样省力、但不解除不可撤回性**的练法：让舞者以极慢的速度全力完成每一个动作。总功率下降，体力消耗下降，但一个慢速的大跳仍然在离地之后不可收回。本文预测这种练法在结构层测验上明显不如标记。

另一边是**同样费力、但保持可撤回**的练法：原速标记，但要求舞者同时做大幅度的手臂动作，使心率上升到接近正式表演的水平。本文预测其结构层收益与普通标记相当。

这一对条件成本不高，值得先做。若结果与预测相反，本文这一节就要重写。

六、为什么它决定这一趟能改哪一层

上一节只是指认了一个被忽略的变量。要让它承重，需要说清它凭什么决定练习的产出。机制分三步，每一步都可以单独被反驳。

第一步：修改需要一个「停」。一段正在展开的动作要被修改，前提是被修改的那样东西能在当下被指认。指认需要把展开中的过程在某一点上截住，让那一刻的状态成为一个可以被指着看的对象。撤回余量为零的动作没有这样的点：从起点到终点它是一个不可分割的整块，可被指认的只有两端。「你这个跳不对」能指的只有落地那一刻的结果。

第二步：可停点的密度决定可修改的粒度。一次练习的修改带宽，可以粗略地写成它每单位时间里能被截停的次数。全力执行把这个数压到接近零——一段八拍的舞句只在句首有一个真正的可停点。标记把它拉到接近连续——任何一拍上都能停。

第三步：可停点的密度必须与要修改的那一层的信息密度相匹配。这是全文的支点，也是最容易滑成循环论证的一步，所以要把信息密度单独定义清楚，不能用「它能不能被停住」来定义它。

本文对信息密度采用一个与可停点无关的测法：**把一段序列按某个粒度打乱，看识别正确率下降多少。**把一段舞句按拍随机交换相邻两拍，熟练舞者立刻认得出不对；把同一段舞句的每一次跳跃在力度上随机上下调整一成，同一批人往往认不出。这说明结构层的信息确实按拍分布，而执行层的信息不是。这个测法只用到识别判断，不涉及任何停或不停，因此可以独立于本文的机制去做。

有了这个独立测法，第三步就成为一条可以被推翻的经验主张：

结构层的信息按拍分布，密度高；执行层的信息按整块分布，一次跳跃只给出一个整体读数。因此要修改结构层，需要高密度的可停点；要修改执行层，需要的恰恰是把整块完整走完。

为什么执行层需要整块？因为执行层的信息只有在整块完成时才存在。半力起跳与全力起跳不是同一个动作的两个强度，它们的发力时序根本不同。

这就得到了本文的承重命题：

标记式练习之所以有效，不是因为负荷被调低，也不是因为难度被调低，而是因为把一组维度归零使动作整段留在了可被随时截停的区间里。**一次练习能修改哪一层内容，由撤回余量决定；与它有多难、多重、多像正式演出无关。**

这条命题三家都不会同意。代表性学习设计不会同意，因为它必须承认存在一类学习在抽掉部分耦合之后反而更好。合意困难不会同意，因为它必须承认存在一个与难度正交的维度，而这个维度上的「更容易」恰恰更有利于学习。认知负荷不会同意，因为它必须承认负荷不是可替换的同一种东西。

但三家的证据合起来只能得出它。舞蹈一侧提供了「削减之后更好」这个必须被解释的事实，运动技能一侧提供了「抽掉信息会改变时序」这条不能被推翻的约束，心理学一侧提供了不可撤回点这把可以把两者分开的尺子。缺任何一家，这一步都推不出来。

七、二维辨别表

只有撤回余量一根轴，得到的仍然是一个名字，而不是一个能用来读数的工具。上一节的第三步已经把第二根轴带了出来：**这一趟练习意图修改的是结构层还是执行层**。两根轴交叉，得到四种练习状态。它们在排练厅、球场与训练房里全都存在，其中两种被反复讨论，另外两种几乎没有名字。

	意图修改结构层 (次序·方位·时值·配合点)	意图修改执行层 (力度·幅度·风险·发力时序)
撤回余量高 (动作随时可收回)	甲·标记 可停点密度与结构信息密度匹配。有效，且是四格中最被低估的一格。舞蹈行业靠经验找到了它，却把它记在「省体力」名下。	丙·空转 看起来在练，实际上要改的那一层的信息根本没有产生。最常见的浪费，且极难被察觉，因为出勤、时长、专注度一切正常。
撤回余量低 (一经发起只能走完)	丁·执行遮蔽 结构缺陷在这一格里既不能被就地修改，也不能被正确记录；它以执行者能力不足的形式显现，并被随后的加练固化。四格中代价最大的一格。	乙·全力演练 执行层的信息只在整块完成时存在，因而这一格不可替代，且必须真做。

读这张表需要注意两点。

第一，四格不是四种价值排序。乙格与甲格同等必要，且不可互换：一个只标记不跳的舞者上不了台，一个只跳不标记的舞者记不住新编舞。丙格与丁格才是问题所在，而它们的问题性质完全不同——丙格是浪费，丁格是错误的知识生产。

第二，判断落在哪一格，不看主观意图的声明，看两个可观测量：这一趟里实际出现的可停点密度，以及事后被记录下来的改动属于哪一层。经常出现的情形是，教练宣称这一趟是「练配合」，实际安排是全速全对抗，事后记录下来的改动是「某某跑动不够积极」。这是丁格的标准签名，而声明与实际之间的落差，不需要任何人不诚实来解释。

八、空转：看起来在练，其实那一层没有出现

丙格值得单独说，因为它是最容易被本文误导出来的一格。若有人从上文读出「可撤回的练习更好」，他会把所有练习都往高撤回余量方向调，而这正是丙格的生产方式。

丙格的典型形态：以远低于正式条件的强度，反复练习一个只有在正式强度下才成立的东西。在举重房里，用空杆反复练抓举的发力时序——空杆的重量不足以产生要被修正的那个时序问题。在球场上，无对抗地反复演练需要在对抗中做出的决策——没有对手的身体存在，那个决策的输入根本不出现。在游泳池边，用橡皮带模拟划水——水的阻力随速度上升，橡皮带的阻力随位移上升，两者的特性相反。

丙格的顽固之处在于它的读数全是正面的。动作看上去越来越标准，主观感觉越来越顺，重复次数在账上，教练的纠正也确实被采纳了。它满足了合意困难所警告的一切条件——当下表现改善而学习未发生——但它的成因不在难度，在于要改的那一层的信息在这个撤回余量下不存在。

这给出一条不涉及任何理论词汇的实操判据：

问一句——**这一趟里，我打算修改的那样东西，有没有真的出现过？**

若答案是「它在正式条件下才出现」，那么无论这一趟多认真、多标准、重复多少次，都在丙格。反过来，若要改的是次序、方位、配合点，那么它在标记里出现得比在全力执行里更完整——因为全力执行时，这些信息被埋在一个只在终点给读数的整块里。

丙格与甲格的外形几乎相同：都是低强度、都能随时停、都在重复。分开它们的不是这一趟的样子，而是要改的那一层的信息密度是否与可停点密度匹配。这也是本文的一个直接后果：**同一个降低强度的动作，在一种意图下是最有效的练习，在另一种意图下是纯粹的空转，而从外面看它们一模一样。**

九、执行遮蔽

丁格是代价最大的一格，也是承重推论所在。

设想一段刚编好的群舞。编导要求六名舞者在一个半拍内完成一次转身并接上下一个方位，而这个转身在物理上需要一拍。这是一个结构缺陷——它写在编舞里，与任何一个舞者的能力无关。现在把这段舞放进全力排练：六个人全速跳，转身接不上，队形散了。

这一遍产生的读数是什么？只有一个：没接上。而「编排要求半拍完成需要一拍的動作」与「这六个人转得不够快」这两个原因，在这一个读数上完全同形。要把它们分开，至少需要两次系统性不同的观察：一次是结构不变而执行者变，看是否仍然接不上；一次是执行者不变而结构变，看把半拍改成一拍是否就接上了。

而全力排练的常规做法恰恰是：**保持结构不变，反复更换执行状态**。再来一遍、注意力集中一点、这一遍加快、明天加练转身。这个安排在信息上只能识别执行差异，它对结构差异完全不敏感——不是因为观察者迟钝，而是因为这个设计的自变量里根本没有结构这一项。

由此得到承重推论：

在撤回余量极低的区间里练结构，结构缺陷不会以结构缺陷的形式显现。它会以执行者能力不足的形式显现，并被随后的加练所固化。

本文称之为**执行遮蔽**。有三点需要说清，否则它会被读成一个已经被说过很多遍的道理。

第一，它不需要任何怪罪动机。下一节将说明，至少有四支文献早已描述过「结构问题被记到人头上」这个现象，而它们给出的原因分别是防御性心理、组织惯例、公正感与统计误读。这些诊断都是对的，它们解释的是**为什么人们愿意这样归因**。本文说的是另一件事：在撤回余量极低的区间里，一个毫无怪罪动机、只看数据、并且真心想找到系统缺陷的观察者，也会得到同一个结论——因为可用的数据里没有能把两者分开的那一项。这是信息上的不可分辨，不是态度上的偏向。

第二，误读的方向是固定的，而且它会自我掩盖。「加练执行」这个处方在短期内几乎总是有效。舞者确实会转得快一点，球员确实会跑得积极一点，于是缺陷读数下降。下降的那部分被记作训练有效，剩下的被记作还需要更多训练。结构缺陷因此从不进入议程，并且它的存在被它自己引发的加练部分补偿掉了。这形成一个不容易被打断的循环：**结构缺陷越是被执行力补偿，它越是显得不存在。**

由此得到一条反直觉的推论：**一个组织的执行力越强，它的结构缺陷越隐蔽。**因为补偿越充分，终点读数越正常，越没有理由回头看结构。这条推论是可裁决的：在规则与任务相同的两个单位之间，本文预测执行力更强的一方，其正式记录中被归为「流程或规范缺陷」的比例更低，而不是更高。

这里要防一个误用。这条推论不意味着执行力是坏事，也不意味着应当放松执行。它只意味着：**执行力与结构诊断能力是两件必须分别建设的事，而前者的成功会系统性地削弱后者的紧迫感。**一个执行力强的组织若不刻意结构诊断另设区间，它会越来越难发现自己的结构问题，而这个过程中的每一步在当时看都是合理的。

还有一个更安静的后果。在丁格里被反复要求「再快一点」的那六个舞者，不只是被误判了，他们还在被训练成一种特定的人：习惯于把结构上做不到的事当成自己没做到的人。这种习得在个体一侧是一种能力

——在不合理的要求下仍然尽力；在系统一侧则是一层越来越厚的缓冲，它使结构缺陷的显现进一步延后。一个组织里若这样的人足够多，它的结构缺陷可以在很长时间内完全不产生任何读数，直到某个无法被个人努力补偿的临界点上一次性显现。

第三，丁格不是「不该全力排练」。全力排练是乙格，不可替代。丁格的定义里有「意图修改结构层」这一项。问题不在强度，在于在一个只能给出执行层读数的区间里，去追一个结构层的问题。相应的处方也不是降低强度，而是把结构层的检验挪到高撤回余量的区间里先做完，再把全力排练用于它真正能改的那一层。

十、四个领域早已看见同一个形状，而它们各自归错了原因

写到这里必须承认一件事：本文所说的形状——**能不能改，决定了改得动哪一层**——并不是第一次被人看见。至少有四个领域各自独立地看见过它，而且看得很清楚。有意思的是，四家给出的原因互不相同，且没有一家的原因在舞蹈标记这个例子里成立。这一节把四家并排放着，因为它们合起来构成了本文能够存在的全部理由。

第一家把原因归给后果。发展心理学关于游戏与漫长未成熟期的经典论述主张：幼体之所以能通过游戏学习，是因为在游戏里错误的后果被削减，探索因此变得可以承担。这条解释覆盖面极广，也确实解释了很多事。但它在标记这个例子上落空：排练厅里的一个全力大跳，后果同样为零——没有观众，不计分，跳砸了没有任何代价——而它照样不可撤回。**后果与可撤回性在这里被干净地分离了：同一个零后果的房间里，全力跳与标记的差别依然存在。**这正是本文可以被检验的地方，也是与这一家的判决性分歧：按后果解释，同一房间里的两种练法应当等效；按本文，不等效。

第二家把原因归给媒介的可逆性。专业实践研究里有一个著名的说法：设计师在草图与推演构成的「虚拟世界」里工作，每一笔都是一次实验，而它之所以能当实验，正因为可以撤销、代价很低。同一传统在设计工具的层面上被写成了准则：让动作可撤销，系统才容许探索。这一家离本文最近，也最容易把本文吞掉。分离点在于：**它谈的是媒介的性质，本文谈的是动作本身的物理。**草图之所以可撤销，是因为纸和铅笔可以擦；大跳之所以不可撤销，是因为人已经离地了。没有任何媒介设计能让一个真实的跳跃在空中被撤销。**这一家的处方是换媒介，本文的处方是在同一个媒介里改动作。**判决性分歧：给一个团队一套可以随时回滚的工具，而其排练动作本身仍然全力全速，这一家预测结构缺陷的识别率上升，本文预测几乎不变。

第三家把原因归给作品的完成度。设计研究里有一组反复被复现的观察。低保真的、看起来还没做完的原型，会引出关于结构与概念的批评。高保真的、看起来已经做完的原型只引出细节反馈，并抑制根本性的意见——人们不太愿意评论看上去已经定稿的东西。这一家在结论上与本文几乎同形，也因此最值得认真对待。分离点有两处。其一，**它的自变量是作品的外观完成度，机制是评论者的社会性犹豫；**本文的自变量是动作的可收回性，机制是数据里有没有可供分辨的那一项。其二，标记并不是低保真的：一段标记保留了全长、全序列、全部配合点，在结构层上它比全力排练更完整，只是在力度上归零。**它不是一份粗略**

的草稿，是一份在某些维度上极其精确、在另一些维度上全空的稿子。判决性分歧：把一段编舞做成一份「粗略但一气到底、不能中途叫停」的呈现，这一家预测会引出结构性批评，本文预测不会。

第四家把原因归给人和组织的动机。这一支最厚。安全科学区分归咎个人与归咎系统两条进路，指出前者顽固是因为它在心理上省事、在制度上方便。组织学习区分单环与双环，指出组织默认在既有规范内修行为、不问支配性变量，而防御性惯例使这个默认自我强化。质量管理那一支更早：一个只靠机械抽样、完全不含技能成分的演示表明，只看排名数据的管理者照样会把系统决定的变异读成工人之间的差异。

最后这一个演示离本文最近，因为它同样不需要恶意。但它的自变量是**统计变异**——数据里有噪声而管理者按点读它。本文的自变量是**识别**——数据里根本没有能把两个原因分开的那一项，与噪声大小无关。即使把噪声降到零，六个舞者每一次都精确地在同一处接不上，那个读数仍然无法区分「编排差半拍」与「他们慢半拍」。**降噪解决不了不可识别。**

把这四家并排看，本文的位置就清楚了。**四个领域各自独立地发现了同一个形状，并各自把它归给一个在舞蹈标记这个例子里根本不存在的原因：这里没有后果，没有媒介可换，作品不粗糙，也没有人有动机怪谁。**剩下的那个变量——动作发起之后还能不能被收回——是四家同时漏掉的一个，而它恰好是唯一一个已经有成熟测量范式的。本文的全部主张，可以压成这一句。

还有一种更紧缩的读法值得先说出来，因为它是本文最尖锐的反对意见：**执行遮蔽在形式上就是一个识别问题**——一个只让执行变化、不让结构变化的观测设计，当然无法识别结构效应，这是方法学常识。这个反对是对的，而本文的回答是：常识部分不在于「不可识别」，而在于**撤回余量逼出了这个设计**。没有人选择只变执行不变结构；是因为在一个一经发起就必须走完的区间里，能被反复调整的只有执行状态，所以那个设计是被区间的物理性质规定的，不是被谁的疏忽规定的。

十一、与最近的几种说法的分界

上一节处理了四个把同一形状归给别的原因的领域。这一节处理另一类相邻者：它们讲的是练习本身，容易与本文混为一谈。每一条给出它自己的最强形态、分界，以及一个在具体情境里会做出相反判断的对照。

部分任务训练与任务简化。把复杂任务分解、简化或分段，何时有利何时有害，取决于任务成分之间的组织程度与相互依赖。它的自变量是任务被切成几块、被简化了多少。分界在于：标记既不切块也不简化，它保留全长、全序列、全部配合点，只把一组维度归零。在这个坐标里，标记没有位置可放。对照预测：取两种在切分与简化程度上完全相同的练习，只让其中一种保持可随时截停。同一段序列，一种以慢速全力做，一种以正常速度轻做。这一支预测两者等效，本文预测后者在结构层测验上更好、在执行层测验上更差。

合意困难与挑战点。这两支都把练习安排映射到一条难度轴上，最优点由学习者能力与任务名义难度共同决定。分界在于：本文主张的维度与难度正交。对照预测：构造两种名义难度相同、撤回余量不同的练习，并在两类测验上分别评分。难度轴预测两类测验上排序一致，本文预测**排序相反**——测次序保持，前

者更好；测落地稳定性，后者更好。同一对条件在两种测验上给出相反的排序，这是难度轴无法产生的预测。

代表性学习设计与练习特异性。两者共同断言练习条件与正式条件的信息结构越接近越好。分界在于：本文主张这条原则对执行层成立，对结构层不成立，而且在结构层上方向相反。对照预测：给两组学习同一段新编排，一组在完全代表性的条件下练，一组在标记条件下练，随后不测表演质量，而是问他们这段编排里哪里有问题。这一支不对这个问题作预测；本文预测标记组能指出编排中的结构性障碍，而全力组更多地报告自己的能力不足。

心理演练与运动想象。想象练习的效应已被反复确认，其理论基础是想象与执行共享表征。标记常被顺手归入「带动作的想象」。分界在于：想象的自变量是**是否产生外显动作**，本文的自变量是**外显动作能否被收回**。纯想象在本文的坐标里是一个特例：撤回余量为满，但没有任何执行层信息。对照预测：在结构层测验上，标记应当优于纯想象，因为标记保留了真实的空间路径与真实的配合点，而这些正是结构层信息的载体；在执行层测验上两者应当同样无效。若实测发现纯想象与标记在结构层上等效，本文关于「保留哪些维度」的说明就需要重写。

刻意练习的反馈条件。这一支把即时反馈与重复修正列为构成条件。它与本文最容易被读成同一件事，因为两者都在强调「改得了」。分界在于反馈的落点：它要求的是**跨次的反馈**——做完一次，得到评价，下一次改；本文说的是**次内的可撤回**——这一次做到一半，就地停住、退回半拍、重来这一处。对照预测：安排一组全力执行并配备逐次即时反馈的练习，与一组标记但完全不给外部反馈的练习。前者按刻意练习应当在结构层上明显更好；本文预测后者不劣于前者，甚至更好，因为结构层的错误需要在错误发生的那一拍上被指认，而跨次反馈只能指认整块的结果。

慢练与分解。音乐与舞蹈教学里最古老的两条经验做法。慢练在本文的坐标里通常提高撤回余量，因而与标记有部分重叠，但两者并不等同，这一点在实践中很重要：慢速全力做一个大跳，撤回余量仍然很低——起跳后仍然收不回；而正常速度的标记撤回余量很高。分界因此可以用一句话说清：**决定的不是慢，是能不能收回**。对照预测：把同一段舞句以「半速全力」与「原速标记」两种方式练习，慢练解释预测前者更利于精确性，本文预测在结构层测验上后者更好，因为前者的可停点密度并未随速度下降而上升——一个大跳在半速下仍然是一个整块。

两条站内相邻。本栏此前有两篇与本文形状接近，须分开。其一讨论一个人做得最好时意识应该在哪里，它撤销的是「意识介入是一个量」，问的是下一个动作由谁定。本文问的不是由谁定，而是定了之后还能不能收回。判决性分歧：一个由教练逐步指定、但每一步都可撤回的练习，那一篇判为接管损耗，本文判为结构层学习的最佳条件。其二讨论一次重写需要「所有使用者同时不在使用」的时刻，那是系统层面的替换机会；本文说的是单次动作发起之后的收回。一个永远等不到那种时刻的系统，其成员仍然可以有大量可撤回的个人练习，反之亦然。

十二、逐领域兑现

一条判断若只在它出生的地方成立，它多半只是那个地方的一句行话。下面六处不是类比，每一处都要求能据此对一件具体的事作出预测。之所以只列六处而不是更多，是因为余下的那些在上一节已经由别的领域先讲过了——重复一遍不是跨域，是再推导一遍。

一、舞蹈排练。预测：新编排的头两遍若以全力进行，编导事后修改编排本身的比例低于以标记进行的情形；且全力组的修改更多落在「要求舞者做到」，标记组的修改更多落在「改动编排」。这是原始情境，也是最容易验证的一处。

二、球类战术演练。预测：新战术在全速全对抗下试跑，事后复盘中被归为「跑位不到位、执行不坚决」的条目比例，显著高于同一战术在慢速无对抗下试跑；后者会更多产生「这个跑位与那个跑位在时间上冲突」这类结构性条目。这条可以在职业队的既有训练记录里做，不需要新实验。

三、体能训练。预测：以远低于目标负荷的重量反复练习发力时序，不会改善目标负荷下的发力时序。这一条在举重界早有经验共识，本文给出的是它的位置：不是「重量不够、刺激不够」，而是要修改的那一层信息在那个负荷下不产生。

四、外科训练。预测：模拟器上的可暂停、可回退训练，对手术步骤次序、器械与人员配置、沟通协议这一层的改进最大，对术中判断与不可逆操作的稳定性改进最小。且对同一批错误，模拟训练的事后记录更多产生流程修订，真实手术的事后记录更多产生个人再培训。

五、航空训练。全动模拟机可以冻结、回退、重置，实机不能。预测：同一支机组的模拟训练记录里，程序与检查单的修订建议密度高于实机训练记录；而实机训练记录里对个人操作技术的评语密度更高。这一预测可以在既有的训练档案上直接检验。

六、应急演练。桌面推演的撤回余量接近满值，全实景演练接近零。预测：桌面推演的事后报告更多产出计划文本的修订，全实景演练的事后报告更多产出人员训练与设备采购条目。下一节将说明，这一处的实测证据同时给本文加上了一条重要约束。

十三、两条让步

第一条：有一类结构缺陷只在不可撤回的区间里存在

应急管理领域提供了一份不太客气的实测。一项覆盖十六家医院与一个一级创伤中心的全实景区域演练研究报告说，桌面推演不足以暴露运作与后勤层面的缺口。而全实景演练暴露出的最普遍缺口是通讯，其次是洗消能力，再次是人手不足与指挥结构不顺。

这份读数直接顶在本文的主张上。按第六节的机制，结构层信息按拍分布、可以在高撤回余量的区间里被完整读出；而这里的证据说，恰恰是最典型的结构性缺陷——通讯链路、指挥层级、人手配置——在桌面推演里读不出来，只在全实景里才出现。

这条约束必须被接受，并且要求把命题改窄：

结构缺陷分两类。一类是**规格类**——次序、职责、时值、接口语义上的歧义与冲突。它们的存在不依赖执行强度，因而可以在高撤回余量的区间里被完整读出。另一类是**载荷类**——只有在全部参与者同时以正式强度运转时才出现的耦合：信道拥塞、资源竞争、注意力饱和、连锁排队。它们在低强度下根本不存在，不是被忽略了，是没有发生。

于是本文原先「结构层在高撤回余量区间里学得更好」这句话被削掉一半：只对规格类成立。对载荷类，唯一的观察窗口就是不可撤回的区间，而那里正是执行遮蔽最严重的地方。

这就留下一个本文解决不了的洞，而且不打算把它填上：**载荷类结构缺陷只在一个系统性地把结构缺陷读成人员缺陷的区间里显现**。已知的补偿手段都不彻底。事后复盘可以补一部分，但复盘的数据仍然来自那个区间；重复多次全实景演练可以让某些缺陷反复出现从而被识别为系统性的，但成本高到大多数组织做不了几次。

本文能提供的只有一条弱的操作建议：在全实景演练里预先安置若干**可叫停点**——预设的暂停时刻，全体停在原地、记录当前状态、再继续——从而在一个整体不可撤回的过程里人为造出几个局部的可撤回边界。这个做法的有效性本身尚未被检验，本文把它列为待检验项，不作断言。

顺带指出另一份读数的复杂性。一项比较桌面推演与全实景演练训练效果的准实验研究报告两者在受训者自评的学习、信任与有用性上没有显著差异，桌面组的均值甚至略高。这份读数不支持「全实景更好」这个一般结论，但也不支持本文——因为它测的是自评，而自评正是本文认为最不可靠的一类指标。在丁格里，受训者的主观体验往往是充实且有收获的，因为加练执行确实带来了可感的改善。这份研究的存在提醒我们：**本文的预测必须落在事后记录的归因分布上，而不是落在参与者的感受上。**

第二条：只在可撤回区间里练，会有一样东西不生成

第二条约束来自外科模拟与表演训练两侧的实践共识。一个只在模拟器上练过的操作者，与一个在真实手术台上练过的操作者，即使在技术动作上难分高下，仍然存在一项差别：**在「这一刀下去就收不回来」这个状态下继续保持稳定的能力**。这项能力的养成条件恰恰是不可撤回本身。可撤回的区间无法生成它，因为在那里「收不回来」从未发生。

同样的事在表演里更直白：排练厅里再完整的标记，也不生成「台上出错了不能停」这一项。舞者对此有一个专门的说法——上过台和没上过台的差别。

于是本文必须补上第二条限定：

可撤回区间只能修改规格类结构，不能生成不可撤回状态下的稳定性。后者只在不可撤回区间里生成，且不可通过任何数量的可撤回练习积累而来。

这条限定与第一条约束合起来，把命题的形状改了。它不再是「高撤回余量更好」，而是**两个区间各自生成一类不可互相替代的东西，而现行的练习设计习惯于用其中一个去做另一个的事**。四格表因此不是一张优劣表，而是一张配对表：先问要生成什么，再选区间。

这一条也堵掉了本文最容易被误用的一条路径。若有人读到「标记优于全力」就把训练全面推向低强度，他会同时得到丙格的浪费与不可撤回稳定性的缺失，而后者的缺失在训练期完全不可见——只在正式场合第一次出错时显现。

十四、怎么测，与那个只需一间排练厅的检验

本文的可用性取决于撤回余量能不能被量出来。下面三条测法互相独立，各有各的成本与局限，可以互相校验。

读数一：直接测不可撤回点。用预期反应抑制范式，在实际动作上测。做法是让受试者按音乐执行某个动作，在随机的一部分试次里于动作展开的不同时刻给出叫停信号，用阶梯法逼近抑制成功率一半的时点。这套方法已经在篮球跳投上做过手指抬起版本，随后又做了整体起跳版本，说明它可以带出实验室、进入接近真实的运动情境。把它用在同一段舞句的两种练法上，直接得到两个读数。成本最高，结果最硬。

读数二：数可停点密度。不测生理，只数行为：在一段录像里，标出所有实际发生的中断——舞者自己停下、编导喊停、动作被改到一半、退回某一拍重来——以每分钟中断次数计。这条读数可以在既有的排练录像上回溯地做，不需要任何设备。坏处是它测的是实际发生的中断，而不是可能的中断，两者在低撤回余量的场合会重合，在高撤回余量的场合可能低估。

读数三：事后记录的归因分布。这是检验执行遮蔽的主要读数，也是成本最低、样本最容易获得的一条。取一批事后记录——排练笔记、训练复盘、演练后报告、术后讨论、发布后复盘——把每一条改进项分类为「改结构」与「改人」，计算前者占比。本文预测这个占比与该场合的撤回余量正相关，且这一相关在控制了事件严重程度、组织层级与报告制度之后仍然存在。

三条读数分开做的理由是它们可能给出不一致的结果，而不一致本身有信息。例如若读数一显示两种练法的撤回余量差别很大，读数三却显示归因分布没有差别，那么执行遮蔽这一步就有问题，而承重命题的前半段仍可能成立。本文希望这两半能被分别裁决，而不是捆在一起接受或拒绝。

另需说明第六节那条信息密度的测法，因为它是全文唯一一处可能滑成循环的地方。做法是：把同一段材料按两种粒度扰动——按拍交换相邻两拍，与在每一次跳跃上随机调整一成力度——请同一批熟练者判断「这一遍与原来一样吗」。两种扰动在物理量上的幅度可以事先配平。若结构扰动的识别率显著高于执行扰动，说明结构层的信息确实按拍分布。这个测法只用识别判断，完全不涉及停或不停，因此不依赖本文要论证的那个机制。

十五、反噬预言与适用边界

任何一条会被写进制度的判断，都应当先预言自己被规避的方式。

若撤回余量进入训练与演练的评估表，最省事的达标法有两条，而且都会使情况变糟。

第一条是**把可撤回的练习当成指标来刷**。低强度、可随时叫停的练习成本低、风险小、场地要求低。一旦它被写进考核，训练计划会迅速向它倾斜。结果是丙格大规模生产：看起来撤回余量很高，实际上要改的那一层信息从未产生；同时不可撤回状态下的稳定性不再生成，而这项缺失在训练期完全不可见。

第二条是**把可叫停点做成文书**。上一节提出的那条弱建议极易退化成一份写在计划里、演练当天走个过场的清单。因为叫停这个动作本身就发生在最容易被记录的那一层，一份漂亮的叫停点设计文件可以在完全不改变演练实质的情况下拿到满分。

由此得到一条唯一不容易被文书化的动作，本文推荐它，且只推荐它：

在一次全力演练或正式发布结束后的第若干周，去数一数事后记录里有多少条改的是结构、多少条改的是人。不问参与者感觉如何，不看计划文件写了什么，只数这个比例。

这个比例难以伪造。要提高它，唯一的办法是真的去改结构；而只要结构没改，改人的条目就会自然地占满记录，无论演练设计得多漂亮。

适用边界有三处。

边界一：只有一个人、且动作本身就是目的的场合。独舞的自我练习、独自的乐器练习里，结构层与执行层的分离仍然成立，但归因错置这一半失去对象——没有第二个人去把结构缺陷记到某人头上。执行遮蔽在这类场合退化为一种自我误判，仍可能发生，但强度小得多，因为练习者同时掌握两侧的信息。

边界二：结构本身就是要被执行强度检验的场合。竞技比赛、真实手术、真实事故处置。在这些场合，人们要的不是「知道结构有没有问题」，而是「在这个结构下把事情做成」。本文对这些场合没有处方，只有一句提醒：它们产生的记录不适合用来诊断结构。

边界三：撤回余量由外部规则规定而非由动作物理决定的场合。一次可以随时撤单的交易、一次有冷静期的合同，其可撤回性来自规则而非力学。本文的机制在这类场合可能仍然成立，但其中的「可停点密度」需要重新定义，因为这里的停并不发生在动作内部。本文不把这类场合纳入主张范围。

另外还有一个洞要交出去。**状态与说法在第三方看来不可分辨**。一个人说「这一遍我在标记」与「这一遍我不想使劲」，从外面看是同一件事。同理，一份事后报告写着「流程需要修订」，可能是真的识别出了结构缺陷，也可能是一种把责任推给流程的常规修辞。本文的第三条读数因此可以被表演。可以做的缓解是取消这条读数的激励性——只在不与考核挂钩的档案上做统计；但只要它被用作评估指标，它就会开始失真。这个洞在本文的框架内无解，因为本文的核心变量本身就是一个可以被声明而难以被外部验证的状态。

十六、本判断对它自己适用吗

按第七节的表，本文自己落在甲格，而且落得毫无悬念：一篇文章从起笔到发表，全程可撤回。任何一句都能删掉重写，任何一节都能退回去改，整篇可以在发表前作废。撤回余量接近满值。

这件事有几重后果，都不利于本文。

第一重：甲格只能修改规格类结构。本文能在这个区间里做完的，只是把概念之间的次序理顺、把歧义削掉、把与相邻说法的分界划清——这些正是规格类的工作。它无法在这个区间里生成任何载荷类的东西：这条判断在真实的排练厅、球场、手术室、演练场里同时运转时会不会遇到本文没有预见的耦合，本文一无所知，而且按自己的主张，也**不可能**在这个区间里知道。

第二重，也是更难堪的一重：本文享受了它所批评的那种豁免。若这条判断是错的，错误会以什么形式显现？按目前的情形，最可能的形式是「没有人去做那六个检验」，也就是以执行的缺席显现，而不是以「这条判断本身有歧义」显现。这恰好是执行遮蔽的镜像：**在一个完全可撤回的区间里提出的主张，其结构缺陷同样不会被正确记录，只不过误读的方向反过来了——不是被记到人头上，而是被记到「尚待检验」这个永远安全的抽屉里。**

第三重：证据分布是对自己最有利的那一种。舞蹈一侧有一个干净的实验结果，运动技能一侧有一套成熟的对立原则，心理学一侧有一把现成的尺子。三样东西凑得如此齐整，本身就应当引起怀疑——一条判断若在证据最容易取得的地方成立，它很可能正是被那些证据的可得性塑造出来的。第十三节里应急管理那份读数之所以重要，正因为它是唯一一处证据不听话的地方，而它确实迫使本文把主张削掉了一半。若有更多这类不听话的读数进来，本文预期还会被削。

第四重，属于工序而非内容。本文的第十节承认，四个领域早已看见同一个形状。这一节是在写作过程中被迫加上去的：最初的稿子里没有它，是把语料往外扩了两轮之后，那四家才一个个跳出来。这本身就是一个读数——**一条判断在它自己熟悉的语料里显得多新，取决于那片语料有多窄。**本文无法保证第三轮扩展不会再跳出第五家。能做的只是把已经跳出来的四家全部写在这里，而不是把它们留在草稿里。

十七、可以让本文失败的六件事，与一条写死日期的赌注

下面六条各自独立，任何一条成立都会让本文的相应部分作废。第一条足够便宜，读到这里的人当中，有能力在三个月内把它做完的应当不止一位。

第一条，最便宜，且是正向读数。取两组舞者或两组球员，学习同一段新的、含有一处刻意设置的结构障碍的编排或战术——例如一个在物理上需要一拍而编排只给半拍的转身，或两条在时间上必然冲突的跑动路线。一组以全力练三遍，一组以标记练三遍。随后不测表现，只问一个开放问题：**这段编排有什么问题？**把回答分为「指向编排本身」与「指向自己或同伴的能力」两类。本文预测标记组指向编排本身的比例显著更高。若两组无差异，执行遮蔽这一步作废，本文只剩下前半段。这个实验需要一间排练厅、两位教师、一份问卷和一位不知分组的编码者。

第二条，排除替代解释。设置一个把记忆与决策负担几乎全部外包出去的全力条件：舞者全力跳，同时由旁人实时报数、报方位、报进出点，使认知负荷降到与标记相当或更低，而动作的可撤回性完全不变。若这个条件产生的收益接近标记，那么认知负荷解释足以说明现象，本文的核心变量是多余的。

第三条，与难度轴的分离。构造名义难度相同而撤回余量不同的两种练法，并在两类测验上分别评分。本文预测两种练法在两类测验上给出相反的排序。若排序在两类测验上一致，说明这里只有一根轴，本文的第二根轴作废。

第四条，与动机解释的分离。取一批全实景演练的事后报告，比较有与没有强免责与匿名报告制度的两类组织。动机一路预测强免责制度会显著提高「改结构」条目的比例；本文预测这个比例主要由演练本身的撤回余量决定，制度变量的效应小得多。若制度变量的效应显著大于撤回余量，本文关于「不需要怪罪动机」的主张就被推翻，执行遮蔽退化为归因偏差的一个特例。

第五条，对本文最不利的一条。若在同一批人、同一段材料上测出：撤回余量高的练法在规格类结构缺陷的识别上并不优于低的，而只是在动作记忆的保持量上更优，那么本文提出的机制就是错的，正确的解释更可能仍在记忆一侧。本文特意把这条写出来，因为它是最容易发生的一种失败，且不需要任何反常的结果。

第六条，跨领域的一致性。本文预测归因分布与撤回余量的关系在多个领域同向。若在其中两个以上的领域测出方向相反的关系，那么它就不是一条跨领域的结构性关系，而至多是舞蹈与球类里的一个局部经验。

六条的读数各不相同。第一条读开放回答的分类比例，第二条读收益的绝对大小，第三条读两类测验上的排序方向。第四条读两个自变量的效应量之比，第五条读缺陷识别与记忆保持之间的分离，第六条读跨领域的方向一致性。

最后是一条写死日期的赌注，用来防止本文以「尚待检验」的形式无限期存活。

到 2031 年 12 月 31 日，若上面第一条检验已被至少两个独立团队完成，而两组在「指向编排本身」的回答比例上没有可复现的差异，本文关于执行遮蔽的主张作废，作者不再申辩。

若到该日期该检验仍无人完成，则本文自行撤回「可检验」这个形容词——一条五年内没有人愿意花三个月去做一次的检验，其可检验性只是一句话。

结 论

把全文压成一句：

一次练习能修改什么，不由它有多难、多重、多像正式场合决定，而由动作发起之后还能不能被收回决定。规格类的结构只在可撤回的区间里能被读出。不可撤回状态下的稳定性只在不可撤回的区间里能生成。而在不可撤回的区间里去追结构问题，得到的将是一份把结构缺陷写成人员缺陷的记录；这份记录还会因为加练在短期内确实有效而被反复确认。

这条判断的价值不在于它推荐了某一种练法。四格里没有一格是应当被取消的。它的价值在于给出了一个此前不存在的配对判据：**先问这一趟要生成的是哪一类东西，再选区间；两者错配时，练得越认真，错误的知识越牢固。**

本文作废的东西有四样。一条最初的强主张（结构层一律在高撤回余量区间学得更好）被应急管理的实测削成只对规格类成立。一条隐含的推荐（把练习往可撤回方向调）被第二条让步取消。一处循环论证（用可停点定义信息密度）被换成一个独立的识别测法。以及一个原始版本里的错觉——以为这个形状是新的。它不新，新的只是它在这一例里的原因。

注 释

1. 本文使用「结构层」与「执行层」这一对区分时，不预设它们对应不同的神经系统或不同的记忆系统。这里的区分是操作性的：一项内容若其读数按拍分布，归入结构层；若其读数只在整块完成时产生，归入执行层。这一操作定义使得同一项内容在不同任务里可能归属不同——例如「发力时序」在一次单独的起跳里属于执行层，而在一段需要与音乐对齐的连续跳跃里，其起始时刻同时是结构层信息。
2. 撤回余量的定义依赖于动作起止的界定，而这在连续动作里并不平凡。本文暂以呼吸断句或音乐小节为切分单位，这是舞蹈与音乐实践中的现成做法，但它是一个约定，不是一个测量。若有更严格的切分方案，第十四节的读数一需要相应调整。
3. 第十三节引用的两项应急管理研究结论方向不同：一项支持全实景不可替代，一项报告两种演练在自评指标上无显著差异。本文认为两者不冲突——前者测的是暴露出的缺口，后者测的是受训者的自评，而本文正是主张自评在这个问题上不可用。但这一调和是本文的读法，不是两项研究作者的主张。
4. 本文未处理团体规模的影响。直觉上，参与者越多，可撤回区间越难维持——一个人停下来，其他人都得停。这可能意味着撤回余量在大规模活动中天然更低，从而使执行遮蔽在大组织里更严重。这个推论尚未被检验，本文不将其列入主张。
5. 第十节列出的四家中，第三家（作品完成度）与第四家（动机）的部分文献本文只读到综述与二手转述，未逐篇核对原始实验细节。这一节的分界线依据的是它们的核心主张，不依赖具体效应量，因此本文认为这个不足不影响分界的成立；但若有人据此指出某一家的实际主张比本文写的更强，本文接受修正。

参考文献

Bainbridge, L. (1983). Ironies of automation. *Automatica*, 19(6), 775–779.

Bruner, J. S. (1972). Nature and uses of immaturity. *American Psychologist*, 27(8), 687–708.

Buxton, B. (2007). *Sketching User Experiences: Getting the Design Right and the Right Design*. Morgan Kaufmann.

- Driskell, J. E., Copper, C., & Moran, A. (1994). Does mental practice enhance performance? *Journal of Applied Psychology*, 79(4), 481–492.
- Ericsson, K. A., Krampe, R. T., & Tesch-Römer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100(3), 363–406.
- Guadagnoli, M. A., & Lee, T. D. (2004). Challenge point: A framework for conceptualizing the effects of various practice conditions in motor learning. *Journal of Motor Behavior*, 36(2), 212–224.
- Jeannerod, M. (1994). The representing brain: Neural correlates of motor intention and imagery. *Behavioral and Brain Sciences*, 17(2), 187–202.
- Kirsh, D., Caballero, R., & Cuykendall, S. (2012). When doing the wrong thing is right. In *Proceedings of the 34th Annual Conference of the Cognitive Science Society*.
- Logan, G. D. (2015). The point of no return: A fundamental limit on the ability to control thought and action. *Quarterly Journal of Experimental Psychology*, 68(5), 833–857.
- Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review*, 91(3), 295–327.
- Naylor, J. C., & Briggs, G. E. (1963). Effects of task complexity and task organization on the relative efficiency of part and whole training methods. *Journal of Experimental Psychology*, 65(3), 217–224.
- Osman, A., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance*, 12(3), 243–258.
- Pinder, R. A., Davids, K., Renshaw, I., & Araújo, D. (2011). Representative learning design and functionality of research and practice in sport. *Journal of Sport and Exercise Psychology*, 33(1), 146–155.
- Proteau, L. (1992). On the specificity of learning and the role of visual information for movement control. In L. Proteau & D. Elliott (Eds.), *Vision and Motor Control* (pp. 67–103). North-Holland.
- Reason, J. (2000). Human error: Models and management. *BMJ*, 320(7237), 768–770.
- Schön, D. A. (1983). *The Reflective Practitioner: How Professionals Think in Action*. Basic Books.
- Slater-Hammel, A. T. (1960). Reliability, accuracy, and refractoriness of a transit reaction. *Research Quarterly*, 31(2), 217–228.
- Soderstrom, N. C., & Bjork, R. A. (2015). Learning versus performance: An integrative review. *Perspectives on Psychological Science*, 10(2), 176–199.
- Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006). Getting the right design and the design right: Testing many is better than one. In *Proceedings of CHI 2006*.
- Warburton, E. C., Wilson, M., Lynch, M., & Cuykendall, S. (2013). The cognitive benefits of movement reduction: Evidence from dance marking. *Psychological Science*, 24(9), 1732–1739.

Wickemeyer, C., Gldenpenning, I., & Weigelt, M. (2024). Action inhibition in a sport-specific paradigm: Examining the limits of action control in basketball. *Psychological Research*, 88(7), 2121–2137.

Wickemeyer, C., Gldenpenning, I., & Weigelt, M. (2025). The limits of action control for deceptive actions in sports: Response inhibition for the basketball pump fake. *PLOS ONE*, 20(11), e0332823.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors*, 27(3), 267–283.

〔应急管理两条读数〕全实景区域演练与缺口暴露： *Journal of Trauma and Acute Care Surgery* (2012), doi: 10.1097/TA.0b013e318265cbb2；桌面推演与全实景演练的准实验比较： *Medicine* (2024), doi:10.1097/MD.00000000000040748。此两条本文只核到期刊、年份与数字对象标识符，未逐字核对作者全名与卷期页码，特此声明。

〔质量管理一路〕关于「只看排名数据的管理者会把系统决定的变异读成人员差异」的经典演示，见戴明关于系统变异与红珠演示的论述，本文按其通行表述引用，未核原始版次。

〔组织学习一路〕关于单环与双环学习的区分，见阿吉里斯与舍恩的组织学习论述，本文按其通行表述引用。

人机分工声明

选题、三家来源的确定、承重命题的形状与全部文字，由王德生与 Claude 共同完成：Claude 负责文献检索、结构组织与初稿撰写，王德生负责判断的裁定、约束条件的取舍与终稿定夺。文中全部实证读数均来自公开发表文献，未包含本研究自行采集的数据。第十七节所列六项检验均为设计，尚未执行。第十节的四家相邻文献，是初稿完成后两轮语料扩展的产物，扩展过程与扩展结果一并写入第十六节第四重。

本文由三个分属不同学科、且互相冲突的理论体系碰撞而成。全部来源均为站外公开文献，可自行核对。 作者：王德生 + Claude · SDE Universes · 学科通融