

可失败余量

论一次练习被削减之后，还能不能当场出错——以及为什么想象在原理上不会失败

体育科学 × 舞蹈艺术 × 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

在“练习应当多大程度上接近目标动作”这个问题上，三个领域给出了互不相容的答案：运动学习的练习特异性原则说越像越好；舞蹈研究发现以缩略、去力量的方式排练（标记）的效果优于全力排练；而心理练习的元分析显示完全不做动作的想象虽优于不练，却弱于身体练习。把后两项并置，沿“执行削减”这条连续谱得到的效应是非单调的——两端低、中间高，而三家分别预测单调递减、单调递增、单调递减，没有一家预测倒 U。本文提出可失败余量：一次练习被削减之后，仍能当场产生由执行者本人察觉的失败的最低限度。练习必须同时给两样东西——一个可被记住的结构，和一次可能失败的机会；削减同时释放资源与削弱失败可能，因此峰值出现在“刚好还能出错”的那一点。想象之所以弱，不是因为它是弱的执行，而是因为它在原理上不会失败：想象里的动作总是成功的。

这一篇由哪三个领域的争论撞成

三个学科由王德生指定。三边对「练习应当多大程度上接近目标动作」给出互相排斥的答案，而把后两家的数字并置，得到的是一条谁也没预测过的倒 U。三条出处均可点开核对（出版社页对自动抓取常有拦截，人工点击可正常打开）。本篇按国际期刊实证稿体例写作，目标 SSCI 二区，文末附投稿路线与本稿现有的不足。

体育科学 · 练习特异性原则

[A Sensorimotor Basis for Motor Learning: Evidence Indicating Specificity of Practice \(Proteau, Marteniuk & Lévesque, 1992\)](#)

学习与练习条件绑定：练习期间可用的感觉信息、执行强度与情境约束会被写进所获得的控制过程，一旦在测试中被改变，表现即下降。其通俗形态是训练的「专项性」——要在什么速度、负荷、情境下比赛，就在那个条件下训练。按这条原则，任何削减都是代价。

<https://doi.org/10.1080/14640749208401298>

舞蹈艺术 · 标记式排练

[The Cognitive Benefits of Movement Reduction: Evidence From Dance Marking \(Psychological Science, 2013\)](#)

舞者排练时普遍把动作压小、抽掉力量、以手代脚，而保持序列、方位与节拍完整。该研究直接操纵排练方式，发现标记组的后续表演质量高于全力组，并提出具身认知负荷假说：全力执行占用大量身体控制资源，压缩执行即释放资源给记忆与整合。按这条解释，削减得越彻底越好。

<https://pubmed.ncbi.nlm.nih.gov/23842957/>

心理学 · 心理练习与运动想象

[Does mental practice still enhance performance? \(Psychology of Sport and Exercise, 2020; 对 Driskell, Copper & Moran 1994 的 24 年复制与扩展\)](#)

完全不产生外显动作的想象练习能提升表现：原元分析报告身体练习 $d = 0.78$ 、心理练习 $d = 0.53$ ；二十四年后的复制在更严格的发表偏倚校正下得到更小的总效应，并再次确认心理练习优于不练但弱于身体练习。按功能等价假说，执行成分越少效力越低。

<https://doi.org/10.1016/j.psychsport.2020.101672>

目 录

一、引言：三条各自有力、却不能同时为真的结论	4
二、理论背景	6
三、冲突的精确表述	8
四、理论框架：可失败余量	10
五、二维判别表	13
六、假设	15
七、研究方案	15
八、与最近的既有解释的分界	18
九、跨领域兑现	22
十、适用边界、反噬预言与使本文为假的条件	24
十一、实践含义	25
十二、本判断对它自己适用吗	26
十三、投稿路线与本稿现有的不足	26

一、引言：三条各自有力、却不能同时为真的结论

1.1 问题

"要学会一个动作，练习应当在多大程度上像那个动作？"这是运动技能获得研究中最古老、也最具实践后果的问题之一。教练排训练计划、编舞者安排排练、外科教学设计模拟器、康复师制定家庭作业，都在回答它。

三个几乎不互相引用的领域各自给出了有经验支撑的答案，而三个答案不能同时为真。

其一，运动学习中的练习特异性原则。 该原则主张，学习是与练习条件绑定的：练习期间可用的感觉信息、执行强度与情境约束会被写进所获得的控制过程，一旦在测试中被改变，表现即下降（Proteau, 1992; Proteau, Marteniuk, & Lévesque, 1992）。它在体能训练中的通俗形态是"专项性"：要在什么速度、什么负荷、什么情境下比赛，就要在那个速度、负荷与情境下训练。该原则预测：**练习越接近目标动作，迁移越好；任何削减都是代价。**

其二，舞蹈研究中的标记式排练。 舞者在排练时普遍采用一种缩略的执行方式：动作幅度被压小、跳跃变成脚下的示意、力量被抽掉、上肢常以手代身，而序列、方位与节拍保持完整。这一做法长期被理解为节省体力。Warburton、Wilson、Lynch 与 Cuykendall (2013) 以实验操纵排练方式（标记 vs 全力），发现**标记组的后续表演质量更高**，并据此提出具身认知负荷假说：全力执行占用了大量用于身体控制的认知资源，压缩执行即释放资源，使记忆与整合得以进行。该假说预测：**执行削减得越多，认知资源越充裕，学习越好。**

其三，心理学中的心理练习与运动想象。 大量研究表明，完全不产生外显动作的想象练习能提升表现。Driskell、Copper 与 Moran (1994) 的元分析报告身体练习效应量 $d = 0.78$ 、心理练习 $d = 0.53$ ，即**心理练习优于不练，但弱于身体练习**；二十四年后的方法学复制在更严格的发表偏倚校正下得到更小的总效应 ($r = 0.131$)，并再次确认心理练习弱于身体练习。功能等价假说主张想象与执行共享表征与神经底物，因而想象的效力随其"生动度"接近执行而提高。该假说预测：**执行成分越少，效力越低。**

1.2 三条结论合起来是不一致的

把第二项与第三项的证据放在同一条轴上——**执行削减程度**，从"全力执行"（削减 0%）经"标记"（部分削减）到"纯想象"（削减 100%）——得到的效应形状是非单调的：

- 全力执行 < 标记 (Warburton et al., 2013)

- 纯想象 < 身体练习 (Driskell et al., 1994)

即：**两端都不如中间。**

三家的理论各自预测的是单调曲线。特异性原则预测随削减单调下降；具身认知负荷假说预测随削减单调上升；功能等价假说预测随削减单调下降。**没有任何一家预测倒 U。**更尖锐的是，负荷假说若被推到底，会得出“纯想象应当最优”——因为它把身体控制的资源占用完全清零——而这与心理练习的元分析结论直接冲突。

这一冲突此前之所以没有被提出，有一个可指认的原因：**三条证据分属三个不共享刊物、不共享方法、不共享被试群体的文献。**舞蹈的标记研究在认知科学期刊发表，心理练习的元分析在应用心理学期刊，特异性原则在运动控制期刊；三者的实验任务、依变量与被试专长水平都不同，因而“把它们放在同一条轴上”这件事没有一个自然的场合发生。本文所做的第一件事，就是把它们放上去。

1.3 本文的主张

本文提出：上述三家共享一个未被审查的前提——**练习的效力是“执行量”这一单一维度的函数**，分歧只在这个函数的斜率符号。撤销这个前提之后，倒 U 有一个简单的解释：

练习必须同时提供两样东西：一个可被记住的结构，和一次可能失败的机会。执行量的削减同时改变这两样——它释放资源（有利于前者），也削弱失败的可能（不利于后者）。倒 U 的峰值出现在“仍能当场失败”的最低执行量上。

本文把“一次练习在被削减之后仍能当场产生可被执行者本人察觉的失败”的最低限度命名为**可失败余量**（failability margin）。

由此得到本文的核心可裁决主张：

在认知负荷被控制相等的条件下，操纵可失败余量将产生显著的学习差异。 负荷假说预测此时无差异；特异性原则预测差异由执行相似度而非失败可能性决定。这一主张可以由本文第七节的研究二直接裁决。

1.4 本文的四项贡献

按目标期刊的体例，此处明示本文自认的贡献，并标出每一项的把握程度。

贡献一（把握最大）：把三条分处三个文献的结论放到同一条轴上，并指出其形状不可能由任何一家的理论生成。 这一步不依赖任何新构念，只依赖一次重新描述；即使本文的构念被否定，这一形状问题仍然成立，且需要某个回答。

贡献二（把握中等）：提出可失败余量作为解释该形状的第二个机制，并给出它与既有第一机制（认知负荷）的嵌套关系。该构念的价值取决于研究二与研究三，二者尚未执行。

贡献三（把握中等）：给出一个可当场使用的判别表与两个问法，把一个理论区分转成教练与教师可用的工具。其实用价值不完全依赖于贡献二成立——即使机制解释被修正，“这次练习里人还会不会当场知道自己错了”仍是一个有用的问法。

贡献四（把握最小）：把该区分推广到十个领域，其中两处反过来给本文加了约束。 推广部分是推论而非证据，本文在第九节逐处标出了读数与竞争预测，以便它们各自被独立检验或否定。

1.5 本文的结构

第二节梳理三条路径并指认各自止步之处，并说明三家为何从未相遇；第三节把冲突精确表述为一条曲线形状的分歧，并给出三家各自的预测曲线；第四节界定可失败余量并给出三个构成条件与操作化定义；第五节升成二维判别表；第六节列出六条可检验假设；第七节给出三项可直接执行的研究方案，含样本量估算与预注册要点；第八节与七种最近的既有解释逐一划出判决性对照；第九节在十个领域兑现；第十节给出适用边界与两条使本理论为假的条件；第十一节讨论实践含义与一条反噬预言；第十二节是本理论对自身的适用；第十三节给出投稿路线与本稿现有的不足。

二、理论背景

2.1 练习特异性原则

特异性原则的现代形态由 Proteau 及其同事在一系列瞄准任务实验中确立。其核心发现是：在有视觉反馈条件下大量练习的被试，一旦撤去视觉，其表现下降的幅度**大于**从一开始就没有视觉的被试——尽管后者练习期间可用的信息更少。这一结果无法由"更多信息导致更好学习"来解释，而支持另一种图景：**学习者所获得的不是一个通用的动作程序，而是一套与练习期间可得信息集绑定的感觉—运动映射**。改变该信息集即改变了任务。

这一原则在应用领域被广泛接受，并以"训练专项性"的形式进入体能训练教科书：训练刺激应在**速度、力量、动作模式、能量系统与情境**上尽可能接近比赛需求。其实践后果是保守的：任何形式的削减都被视为不得已的妥协，其正当性来自伤病、疲劳或安全，而不来自学习本身。

它停在哪里？ 停在**它把"练习条件"当作一个整体**。特异性原则处理的是练习条件与测试条件之间的匹配，而没有把练习条件本身拆成可分离的成分并追问每一成分各自的作用。标记式排练恰恰是一个成分被系统拆开的现成案例：序列、方位与节拍被完整保留，力量与幅度被抽掉。若特异性原则成立于成分层面，则应当预测"力量与幅度的迁移受损而序列的迁移不受损"；而 Warburton 等人的结果显示整体表现更好——这已经超出了原则的预测范围。

2.2 标记式排练与具身认知负荷假说

Warburton 等人（2013）的贡献有两层。第一层是把一个长期被视为纯粹节能措施的实践，重新指认为**认知策略**：标记不是"省力气地过一遍"，而是"在不被身体控制占满的情况下过一遍"。第二层是他们的实验设计——直接操纵排练方式并测量后续表演质量，得到标记组更优的结果。

具身认知负荷假说的推理链是清楚的：复杂动作的在线控制本身消耗中央资源；排练期间需要同时完成的另一件事（记住序列、整合音乐与空间、理解表现意图）与之竞争同一资源；压缩执行即降低控制需求，从而释放资源给后者。

它停在哪里？ 停在它的预测是单调的。若资源竞争是唯一机制，则削减越彻底、释放越多、学习越好，其极限（零执行）应当最优。心理练习的元分析否证了这个极限。因此负荷假说要么必须承认在某处存在一个反向作用，要么必须主张"想象"不在这条连续谱上——而后者需要独立论证，且与功能等价假说的核心主张（想象与执行共享表征）冲突。本文提供的正是那个反向作用。

需要说明的是，本文并不主张负荷假说是错的。本文主张它是**不完整的**：它正确地刻画了曲线的上升段，而对下降段沉默。

2.3 心理练习与运动想象

心理练习的效力已被数轮元分析确认，同时也被反复确认为**弱于身体练习**。除总效应外，两项调节效应对本文尤为重要：

其一，**专长调节**。Driskell 等人报告，新手在认知性任务上从心理练习中获益更多，而有经验者在认知性与动作性任务上均获益。若想象的效力仅由表征生动度决定，专长的作用应当是连续的；而若如本文所主张，想象的效力取决于**想象中是否包含可被识别的失败**，则专长调节有一个更具体的解释：**只有已经知道"会在哪里出错"的人，其想象中才可能出现失败**。新手想象自己完成动作时，想象的内容是一个成功的、平滑的、没有阻碍的版本——因为他还不知道阻碍在哪里。

其二，**保持间隔调节**。心理练习的效应随练习与测试之间的间隔延长而衰减，比身体练习衰减更快。这与"想象未产生需被修正的差异、因而未引起持久的重新校准"这一读法一致。

它停在哪里？ 停在它把想象的不足归因于表征的"弱"——想象被理解为一次强度较低的执行。本文主张不足的性质不同：想象不是弱的执行，而是**不可失败的执行**。一个可以被无限次成功完成的动作，无论被想象得多么生动，都不产生任何需要被修正的东西。

2.4 三条路径共享的东西

三条路径的自变量分别是"练习与目标的相似度"、"释放的认知资源量"、"表征的生动度"。三者形式不同，但共享一个类型：**它们都是关于"执行量"的单调函数**。分歧只在符号。

正因为共享这个类型，三者无法共同解释一条倒 U 曲线：任何单调函数的组合若要产生倒 U，必须至少引入两个方向相反的机制，而三家各自只提供了一个。本文引入的第二个机制，就是可失败余量。

2.5 三家为什么从未相遇

一个自然的质疑是：这三条结论都不新，为什么它们的不一致此前没有被指出？

原因不在于难，而在于**三条证据分处三个不相交的学术生态**，且每一个生态都有充分的理由不去看另外两个。

其一，刊物不相交。 特异性原则的核心文献发表在运动控制与实验心理学期刊；标记研究发表在综合性认知科学期刊，并被舞蹈研究与具身认知两个小共同体接住；心理练习的元分析发表在应用心理学与运动心理学期刊。三条线的引用网络重叠极小——一个可以被直接核查的断言是：**标记那篇论文的参考文献表里没有任何一条特异性原则的核心文献，而心理练习元分析的更新版里也没有标记研究。** 这一点投稿前应引文数据库逐条核实并在文中给出具体数字。

其二，被试群体不相交。 标记研究以受过训练的舞者对象，心理练习的元分析大量样本是新手与运动员，特异性研究多用未受训的大学生完成实验室瞄准任务。**专长水平的系统差异使三方各自的结论在对方眼里天然可疑，因而不必解释。**

其三，"削减"在三方各有不同的名字。 在体能训练里它叫"降负荷"，是不得已；在舞蹈里它叫"标记"，是行规；在心理学里它叫"想象"，是一种独立的干预手段。**同一条连续谱被三个词切成了三段，而每一段都被当作一件独立的事情来研究。** 这是本文之所以可能的直接原因：把三个词换成一个量（执行削减度 A ），三段自动接成一条曲线，而曲线的形状立刻显出问题。

这三条原因也构成对本文的一个提醒：**本文的贡献首先是一次重新描述，其次才是一个构念。** 若重新描述是错的（例如三者确实不在同一条轴上），构念也就失去了对象。第三节 3.2 处理的正是这一点。

三、冲突的精确表述

3.1 一条轴与三条曲线

定义**执行削减度** $A \in [0, 1]$: $A = 0$ 为全力执行（full out）， $A = 1$ 为完全无外显动作（纯想象），中间值对应不同程度的缩略执行（标记）。令 $L(A)$ 为学习收益。

三家的预测：

理论	预测的 $L(A)$ 形状	极值位置
练习特异性原则	单调递减	$A = 0$ 最优
具身认知负荷假说	单调递增	$A = 1$ 最优
功能等价假说	单调递减	$A = 0$ 最优
现有证据	两端低、中间高	$0 < A^* < 1$

现有证据来自两项独立结果的并置： $L(\text{标记}) > L(0)$ (Warburton et al., 2013) 与 $L(1) < L(0)$ (Driskell et al., 1994)。**任何单调曲线都无法同时容纳这两条不等式。**

3.2 三条可能的辩护及其代价

在提出新构念之前，必须先处理三条可能的辩护——它们都能使冲突消失，但各自要付出明确的代价。

辩护一：想象不在这条轴上。 若主张纯想象是一种与执行削减性质不同的活动，则第二条不等式不构成对负荷假说的约束。**代价：**这与功能等价假说的核心主张冲突，并使"标记"的位置变得无法定义——标记究竟是"弱执行"还是"带一点动作的想象"？该辩护要求给出一条把二者分开的独立判据，而目前没有。

辩护二：任务不同。 舞蹈的编舞记忆与瞄准、投掷等实验室任务在结构上差异极大，二者的结果不可比。**代价：**这使特异性原则与负荷假说各自的适用范围大幅收窄，并且要求解释为何舞蹈任务上"削减更好"、实验室任务上"削减更差"——而这个解释本身就需要一个新的调节变量，即本文所要提供的东西。

辩护三：依变量不同。 表演质量与瞄准误差测的不是同一件事。**代价：**这条辩护最有力，也最容易检验——第七节的研究一在**同一任务、同一依变量**上取五个削减水平，若倒 U 不出现，本文即被驳倒。

本文接受这三条辩护都尚未被排除，并把排除它们作为研究一的首要目的。**在研究一完成之前，本文的地位是"一个由两条独立文献并置所提示的、尚未在同一实验内被观察到的形状，以及一个解释它的构念"，而不是"一个已被确立的效应"。**这一点在第十三节还会再说一次。

3.3 冲突的位置

三家的冲突全部发生在同一个位置：**当执行被削减时，究竟丢失了什么。**

特异性说丢失的是"与目标条件的一致性"；负荷说丢失的是"无关紧要的东西，反而卸掉了包袱"；功能等价说丢失的是"表征的强度"。

三者共享的前提是：**丢失的东西只有一样，因而效应必然单调。**

本文撤销这个前提，主张削减同时改变两样彼此独立的东西：**释放的资源**（随削减单调上升）与**当场失败的可能性**（随削减单调下降）。学习收益是二者的乘积形式，因而必然呈倒 U——而峰值的位置由后者何时归零决定。

3.4 三家各自为何到不了这一条

三家都不缺能力，缺的是各自研究对象的结构性条件。

特异性原则到不了，因为它的实验范式里"失败"永远存在。 瞄准、投掷、追踪这类实验室任务本身就带着一个不可移除的外部约束——靶。无论练习条件怎么变，被试总能当场知道自己偏了多少。**在一个可失败性恒为正的范式里，可失败性不可能成为一个变量。** 它因此只能被吸收进背景，而所有解释的重量落到剩下的那个变量（信息集的一致性）上。

标记研究到不了，因为它只有两个水平。 Warburton 等人的设计是标记与全力的二水平对比，这足以证明"削减不总是有害"，但**两点连不出曲线**。要看到倒 U 至少需要三个水平，而第三个水平（纯想象）在

舞蹈排练的现实中不是一个可比的选项——没有编舞者会让舞者闭眼想象一遍来代替排练。**该领域的实践本身排除了那个能暴露倒 U 的水平。**

心理练习研究到不了，因为它的对照组是"不练"。心理练习文献的标准设计是"想象组 vs 身体练习组 vs 不练组"，其关心的是"想象有没有用"。在这个问题框架里，想象的位置是**从零往上看**（比不练好多少），而不是**从执行往下看**（比削减执行差多少）。**参照点的选择决定了看得见什么：**从零往上看，想象是一项成就；从执行往下走，想象是曲线的塌陷端。该领域从未把"部分削减"放进设计，因而不曾看到中间那一段。

三家的边界因此都不是认识能力问题：一个的范式里那个变量恒为正、一个的实践排除了关键水平、一个的参照点方向相反。三者并排放着，那个共同的变量才显出来。

四、理论框架：可失败余量

4.1 定义

可失败余量（failability margin, F）：一次练习在其执行被削减之后，仍能**当场**产生一个**由执行者本人在动作进行中即可察觉**的失败的最低执行限度。

三个限定词各自承担一件事，缺一即不成立：

- **当场**——失败必须在动作进行中或紧随其后发生，而不是事后由录像、教练或成绩单告知。事后告知的是外部反馈，属于另一族机制（见第八节与刻意练习的分界）。
- **由执行者本人察觉**——不需要任何外部裁判。这一条把本构念与"反馈"整族分开：反馈是被给予的，失败是被撞上的。
- **最低执行限度**——它是一个阈，不是一个量。可失败余量为正意味着"还能出错"，为零意味着"无论怎么做都不算错"。

4.2 三个构成条件

一次练习要具有正的可失败余量，需同时满足：

其一，存在一个不由执行者设定的约束。节拍、同伴的位置、器械的边界、地面的摩擦、重力、他人的动作。若全部约束都由执行者自己在心里设定，则任何偏离都可以被同时重设为"我本来就想这样"——**失败在定义上不可能发生。**这正是纯想象的处境：想象中的节拍由想象者自己维持，想象中的同伴不会站错地方。

其二，存在一个可被当场读出的差异信号。撞到、踩空、错拍、够不着、失衡。这一条要求削减保留了足以让约束咬合的那部分执行：手代脚的标记仍然占据空间与时间，因而仍然可能与同伴的路径冲突、与音乐的节拍错开。

其三，失败的代价足够低以致可以被容许发生。 若一次失败意味着受伤或不可挽回的损失，执行者会预先降低难度以规避它，从而在实际上把余量重新压到零。这一条解释了高风险任务上练习设计的困难，也给出一个非平凡的推论：**安全不是可失败性的敌人，它是可失败性的前提**——只有当失败便宜，人才允许自己失败。

4.3 为什么形状是倒 U

设削减度 $A \in [0, 1]$ 。两个方向相反的量随 A 变化：

- **可用资源** $R(A)$ ：随 A 单调不减。这是负荷假说所刻画的部分。
- **可失败余量** $F(A)$ ：随 A 单调不增，且在某个 A_0 处降为零——一旦执行被削减到不再与任何外部约束咬合，失败即不可能。

学习收益近似为二者的合取而非叠加：

$$L(A) \approx g[R(A)] \cdot \mathbf{1}[F(A) > 0] + c$$

即：资源越多越好，**但当仍可能失败时**；一旦 F 归零，收益跌至一个仅由表征复述提供的低平台 c （这正是心理练习相对于不练的那部分正效应）。

由此得到三条形状预测：

1. $L(A)$ 在 $A \in [0, A_0)$ 上递增（负荷说的上升段）；
2. 在 A_0 处出现**不连续或急剧的下降**，而非平滑衰减；
3. 峰值 A^* 紧邻 A_0 之左，即**最优练习是"刚好还能出错的最省力版本"**。

第二条是本文与所有平滑单调模型最容易分开的地方：本文预测的不是一个缓坡，而是一道**坎**——坎的位置由约束是否还咬合决定，与能耗、幅度等连续量只有间接关系。

4.4 与"认知负荷"解释的关系与区分

本文与负荷假说不是竞争关系而是**嵌套关系**：负荷假说是本文在 $F > 0$ 区间内的特例。二者在两个位置上给出不同预测：

位置一（关键实验）：令两组的执行幅度、能耗与主观负荷相等，仅操纵是否保留一个外部约束（例如：甲组标记时须与同伴保持队形并跟随外部节拍；乙组标记时单独进行、自定节拍）。**负荷说预测无差异；本文预测甲组显著更优。** 研究二即此设计。

位置二（想象条件下的植入）：在纯想象中人为植入可失败性（随机中断、外部节拍与想象速度不匹配、事后即刻强制报告"哪一处没跟上"）。**负荷说预测无变化或轻微变差（增加了负荷）；本文预测效应量向身体练习靠拢。** 研究三即此设计。

这两处是本理论存在与否的分界线。若两处都无差异，本文应当被并入负荷假说，不必另立名目——这一点在第十节写成了明确的作废条件。

4.5 阈的移动：一条三家都不作的预测

前面几条预测都是关于曲线**形状**的。形状可以被多种模型凑出来，因而不夠硬。本节给出一条关于**峰值位置如何移动**的预测——它比形状更难被凑，也是本文唯一一条与三家全部分开的定量预测。

按 4.3，峰值 A \ 紧邻可失败余量归零处 A_o 。而 A^*_o 由**约束的容差**决定：容差越紧，越轻微的削减就已经不再触发失败；容差越松，要削减得很厉害才不再出错。因此：

收紧约束的容差， A_o 左移，峰值随之左移；放松容差，二者右移。而在此过程中，负荷曲线 $R(A)$ 完全不变——因为容差不改变动作本身要花多少力气。

这条预测可以用一个不改变任何执行参数的操纵来检验。以外部节拍同步为例：把"算作跟上"的时间窗从 ± 120 毫秒收紧到 ± 40 毫秒。被试的动作幅度、速度、能耗、注意负荷都不必变，变的只是**多小的偏差开始算错**。

理论 收紧容差后峰值如何移动

练习特异性原则 不动（相似度未变）

具身认知负荷假说 不动（负荷未变）

功能等价假说 不动（表征强度未变）

功能保真／任务对齐 不动或右移（对齐度未变）

本文 显著左移

四家全部预测不动，只有本文预测移动——这是本文目前最干净的一处分离，因为它不依赖任何关于"资源"或"相似度"的假定，只依赖"失败何时算数"这一个量。

操作上它也很便宜：在研究一的五水平设计里增设一个容差因子（宽窗／窄窗），即成为 5×2 的被试内设计，样本量需求仅小幅上升，而所得的是一条**交互效应**——本文预测容差与削减度的交互显著，且窄窗条件下的峰值出现在更靠近全力执行的一侧。

需要预先说明一处风险：收紧容差同时会提高失败**发生率**，而失败发生率本身可能通过挫折或注意分配影响学习。因此该操纵必须配一个控制条件——**发生率匹配但容差不变**（通过略微加快节拍使错误率与窄窗组相当）。若峰值移动在控制了发生率之后消失，则该预测被拒。

4.6 操作化：怎么测可失败余量

一个构念若不能被测，它只是一个说法。本文给出四种由粗到细的测法，四者可互相校验。

测法一（事件计数，最直接）：在一次练习中记录**当场可见失败事件**的次数——错拍、越线、与同伴冲突、失衡、够不着、被判超时。要求由第三方按预先定义的清单计数，不依赖执行者自陈。**优点**：客观、

可自动化（动作捕捉、压力垫、音频对齐皆可）；**缺点**：它测的是失败的实际发生，而可失败余量说的是失败的**可能性**——一次练习可以余量为正而恰好一次没错。因此本测法在低次数区间不敏感，需与测法二合用。

测法三与测法二的分工在这里必须说清楚。

测法二（扰动探针）：在练习中随机注入一次小扰动（节拍器突然提前 60 毫秒、投影中的同伴位置偏移 15 厘米、地面标线短暂点亮），观察执行者是否在**无人提示的情况下**产生可观察的调整。产生调整即证明约束仍在咬合， $F > 0$ ；毫无反应即 $F = 0$ 。**这是本文推荐的主测法**：它直接测“可能性”而非“发生率”，且探针可以做得很小，不改变练习本身。

测法三（结构清单，最便宜）：逐项核对四件——(a) 是否存在至少一个不由执行者设定的约束；(b) 该约束在本次练习的执行幅度下是否仍会被触发；(c) 触发时执行者能否在 1 秒内察觉；(d) 触发的代价是否低到执行者愿意让它发生。四项全为“是”则 $F > 0$ 。**优点**：教练与教师可当场使用，无需设备；**缺点**：粗，且 (c) 依赖判断。

测法四（自陈校验，用于研究）：练习结束后即刻要求执行者标出“我知道自己错了的那些时刻”，与测法一的第三方计数做一致性比较。**二者的落差本身是一个有价值的读数**：第三方计到而执行者未察觉的次数，是“约束在咬合但人没接住”的量，它预测该次练习的学习收益低于其失败次数所提示的水平。这一落差在新手身上应显著大于专家身上，构成对 10.1 第四条边界的一个直接检验。

四种测法中，研究一与研究二使用测法一与测法四，研究三使用测法一与测法三。**若资源允许，测法二应被加入研究二作为操纵检查**——它能直接证明“去约束组的约束确实不咬合了”，而不是仅仅假定。

五、二维判别表

单一维度（A）不足以刻画实践中的情形，因为同一个执行量可以配上不同的约束条件。第二条轴取**失败可见性**：练习中是否存在当场可察觉的失败（即 $F > 0$ ）。

	失败当场可见 ($F > 0$)	失败不可见 ($F = 0$)
执行量高	格 I 全力排练 。失败可见，但资源被身体控制挤占；有效而昂贵，且疲劳会进一步压缩可用资源	格 III 空转重复 （本文的靶格之一）。又累又学不到：单独一人、无节拍、无对手、无判定标准地把动作做满一百遍
执行量低	格 II 标记（最优格） 。资源被释放而约束仍咬合：与同伴的队形、外部节拍、场地边界都还在起作用	格 IV 安全复述 （本文的靶格之二）。纯想象、默读式过一遍、看视频“在脑子里跟一遍”；感觉良好、成本低、且 永远不会出错

四格中值得单独指认的是**格 III 与格 IV**。

格 III 的危险在于它看起来最像“刻苦”。执行量高，因而主观努力感、疲劳感、时间投入都很高，而由于没有任何外部约束咬合，练习不产生任何需被修正的差异。它在体能训练、乐器练习与技能培训中极为常

见：一个人对着墙练一百次发球、在空房间里过一百遍套路、在没有病人的房间里演练一百次问诊。**它的读数特征是：主观努力高而当场失败次数为零。**

格IV的危险在于它在现代条件下正在急剧扩张。看教学视频、读操作手册、在脑中"过一遍"流程、在会议前默演一遍发言——这些活动的共同特征是在**原理上不会失败**：视频里的人总是做对的，脑中的自己总是流畅的。本文并不主张这些活动无效（第 c 项是正的），而主张**它们的效力有一个由结构决定的天花板，且这个天花板不会因为做得更多、更投入、更"生动"而抬高。**

四格给出两个可当场使用的问法：

1. "上一次练习里，你在什么时刻知道自己错了？" 答不出具体时刻的，F 很可能为零。
2. "如果你今天做得比昨天差，这次练习会有什么不同？" 若答案是"没有不同"，则该练习落在右列。

第二问比第一问更硬，因为它不依赖回忆，而依赖练习结构本身。

5.2 三个看似反例

一个新观念最先要处理的不是支持它的例子，而是看上去反对它的例子。以下三个是最容易被举出的。

反例一：静态拉伸与体能训练。 静态拉伸、有氧基础训练、力量训练中的常规组次，都不存在"当场失败"，而它们显然有效。**回应：**本文的适用对象是**技能获得**，不是体能适应。体能训练的机制是生理适应（肌肉、心肺、结缔组织对负荷的响应），其自变量是负荷与恢复，与是否出错无关。**这一区分不是补丁，它给出一个可检验的边界：**在同一次训练中，若同时包含体能成分与技能成分，可失败性的操纵应只影响技能成分的学习，不影响体能指标的变化。这一预测可在同一批被试上直接检验。

反例二：冥想式练习与太极。 缓慢、无外部约束、以内感受为主的练习传统被广泛认为有效。**回应：**这类练习的外部约束常被误认为不存在，实则存在且极严——**重力与平衡**。缓慢移动时的重心控制随时可能失败，且失败当场可察觉（晃、需要额外用力、无法维持）。**测法二在这里给出一个可做的检验：**在缓慢练习中引入极轻微的支撑（扶墙一指），按本文预测应显著降低学习收益，而按"放松—内感受"解释应无影响或有益。

反例三：观摩学习（observational learning）确实有效。 大量证据表明观看他人示范能促进学习，而观看在结构上不会失败。**回应：**这是三个反例中最有力的一个，本文必须让步。观摩学习的效力被广泛认为由观察—执行匹配系统与认知表征建构承担，它与本文所说的机制是**两条并行的通路，而非同一条**。本文的主张因此必须收窄为：**在包含外显执行的练习内部，可失败余量是决定其效力的关键变量；而不含执行的活动（观摩、阅读、讲解）自有其通路，其效力上限由该通路决定。**

第三条让步的后果不小：它意味着格IV并非一律低效，而是**低效仅就"通过执行来学"这条路而言**。一个把观摩通路用足的安排（结构化观摩、对比示范、错误示范）可能在格IV内达到相当高的效力。**本文不涵盖那条通路，也不主张它次要。**

六、假设

H1（形状） 在同一任务与同一依变量下，学习收益随执行削减度 A 呈倒 U 形，峰值位于 $0 < A^* < 1$ 。

H2（峰值位置） A^* 出现在可失败余量刚要归零的削减水平上，而非主观负荷最低的水平上。操作化：在同一实验中同时测量各水平的当场失败次数与主观负荷（NASA-TLX）， A^* 与“失败次数首次降至零的水平”的距离应显著小于其与“负荷最低水平”的距离。

H3（关键分离） 在执行幅度、能耗与主观负荷统计上相等的两组之间，保留外部约束者的学习收益显著高于未保留者。

H4（想象的可修复性） 在纯想象练习中植入外部约束（外部节拍、随机中断、即刻强制差异报告），其学习效应量显著高于常规想象练习，且与身体练习的差距缩小。

H5（专长调节的重新解释） 心理练习的专长调节效应可由“想象中是否包含失败”中介。操作化：以“想象过程中报告出现困难/卡顿的比例”为中介变量，该变量应显著中介专长与心理练习效应之间的关系。

H6（疲劳的双重作用） 在格 I 内，疲劳的增加同时降低可用资源与降低失败阈（更容易出错）；因此全力排练的学习收益随疲劳呈先升后降，且其峰值早于表现峰值。

H6b（阈的移动） 收紧外部约束的容差将使学习收益曲线的峰值向低削减一侧移动，且该效应在控制失败发生率后仍然存在；负荷曲线在此操纵下不变。

H7（时间进程） 在一个完整的排练或训练周期内，标记（格 II）相对于全力执行（格 I）的优势随目标事件（演出、比赛、考核）临近而单调下降，并在最后阶段消失或反转。这一预测源自 9.1 第二条让步：临近目标时，练习的目的从获得转为稳定，特异性重新占据主导。

H8（落差的专长调节） 第三方计到的当场失败次数与执行者自陈察觉次数之间的落差，随专长上升而缩小；且该落差可负向预测同次练习的学习收益，即使控制失败总次数。

H1、H2 与 H6b 由研究一（增设容差因子后为 5×2 设计）检验；H3 由研究二检验；H4 由研究三检验；H5 与 H8 可由现有心理练习数据与研究一的过程数据的二次分析检验；H6、H7 留待后续，其中 H7 可用现有排练日志回溯检验，成本极低。

七、研究方案

三项研究均按预注册标准撰写，含样本量估算、主要终点、排除标准与分析计划。三项可独立执行；若资源有限，研究二是唯一不可省的一项——它是本理论与负荷假说的唯一分界点。

7.1 研究一：执行削减的五水平连续谱

目的 在同一任务、同一依变量上检验倒 U（H1）与峰值位置（H2）。

设计 被试内五水平，拉丁方平衡。任务为一段 32 拍、包含 12 个动作元素的新编舞蹈序列（改编自 Warburton 等人所用材料的公开描述，并另编等难度的四套以供平衡）。五个削减水平：

水平	操作定义	预期 F
A0 全力	全幅度、全力量、按谱执行	高
A1 轻减	幅度约 70%，跳跃改为踮起	高
A2 标记	幅度约 40%，以手代脚，保持方位与节拍	中
A3 微标记	仅手部示意，站立不移动，仍随外部节拍	低
A4 纯想象	闭眼想象，不产生任何外显动作	零

关键控制 A0-A3 全程播放外部节拍与同伴录像投影（保持外部约束恒定），**唯有 A4 在结构上无法保留同伴约束**——这一不对称本身即本文所主张的机制，将在讨论中明确处理，并由研究三补上。

依变量 主要终点：24 小时后无提示复现的**动作元素正确率**（两名不知分组的舞蹈教师独立评分，ICC 报告）。次要终点：序列顺序错误数、时序偏差（与节拍的均方误差）、表现质量七点量表。

过程测量 每水平练习中记录**当场失败次数**（错拍、与投影队形冲突、重心失衡）；每水平结束即测 NASA-TLX 与主观努力感（Borg CR-10）。

被试与样本量 中级以上舞蹈训练者（≥3 年）。以中等效应（ $f = 0.25$ ）、五水平被试内、 $\alpha = .05$ 、 $power = .80$ 、相关 $r = .5$ 估算，需 $N \approx 34$ ；考虑 15% 脱落与评分者可靠性折损，计划招募 $N = 45$ 。

分析 对 L(A) 拟合线性与二次趋势对比；H1 要求二次项显著且方向为负、线性项不足以解释。H2 以两个距离的自助法置信区间比较。

预注册的证伪条件 若二次趋势不显著、而线性趋势显著（无论方向），H1 被拒。

7.2 研究二：负荷等同、可失败性分离（关键实验）

目的 在负荷相等条件下分离可失败性（H3）。这是本理论与具身认知负荷假说唯一的判决性对照。

设计 被试间两组，均在 A2（标记）水平练习同一序列，练习时长与重复次数相同。

- **甲组（保约束）**：与两名同伴同场标记，需保持三角队形（地面标线，越线即当场可见）；外部节拍器播放，须与之同步。

- **乙组（去约束）**：单独在同样大小的场地标记，无地面标线、无同伴、无外部节拍（自行默数）。

负荷等同的措施与验证 两组动作幅度由同一段示范录像界定并由录像回看核验（幅度差异以关节角度均值检验，要求无显著差异）；能耗以心率与主观努力感核验；认知负荷以 NASA-TLX 六个分量表逐项比较，**要求两组在总分与各分量表上均无显著差异**（等价性检验，TOST，界值取 0.4 SD）。若负荷未能等同，本研究不构成对负荷假说的检验，应作为失败的操纵报告而非阴性结果。

依变量 同研究一主要终点。

样本量 以 $d = 0.6$ （中等偏上，依据 Warburton 等人组间差异的量级）、双尾 $\alpha = .05$ 、 $\text{power} = .80$ 估算，每组 $n = 45$ ，合计 $N = 90$ 。

竞争预测

理论 对甲乙差异的预测

具身认知
负荷假说 无差异（负荷已等同）

练习特异 无差异或乙组略优（乙组更接近"独自表演"？取决于测试情境，需在预注册中固定测试为同性原则 场三人，此时特异性预测甲优——**故本研究不能分开本文与特异性，须与研究三合看**）

本文 甲组显著优，且中介变量为练习中的当场失败次数

中介分析 以练习期间当场失败次数为中介，检验甲乙分组 $\rightarrow$ 失败次数 $\rightarrow$ 学习收益的间接效应（Bootstrap 5000 次）。本文预测间接效应显著且直接效应大幅衰减。

预注册的证伪条件 若在负荷成功等同（TOST 通过）的前提下甲乙无显著差异（且等价性检验支持等价）， H_3 被拒，本文应被并入负荷假说。

7.3 研究三：在想象中植入失败

目的 检验可失败性是可加的（ H_4 ），并回答研究一中 A_4 水平的不对称问题。

设计 被试间四组，练习时长相同：

- **组1 常规想象**：闭眼按自己的节奏想象整段。

- **组2 外节拍想象**：闭眼想象，但须与外部节拍器同步，且节拍器速度略快于被试舒适速度（制造可被察觉的"跟不上"）。

- **组3 中断想象**：闭眼想象，随机在 15%–20% 的时点插入一声提示音，被试须立即报告"当前在第几个动作"（答不出即为一次当场可见的失败）。

- **组4 身体练习（阳性对照）**： A_2 水平标记，保约束。

依变量 同上。另记录组2 的失同步次数与组3 的答不出次数，作为可失败性的操作化读数。

样本量 四组、 $f = 0.25$ 、 $\alpha = .05$ 、 $\text{power} = .80$ ， $N \approx 180$ （每组 45）。

竞争预测 功能等价假说预测组1、2、3 无差异（生动度未被操纵）或组2、3 更差（外部干扰降低想象质量）；负荷假说预测组2、3 更差（增加了额外负荷）；**本文预测组2、3 显著优于组1，且向组4 靠拢。**

预注册的证伪条件 若组2、3 不优于组1（或更差）， H_4 被拒。**这是本文最容易失败的一条，也是我认为最值得先做的一条**——它成本最低（无需舞蹈场地与同伴协调），且它若失败，本文的核心机制（失败而非资源决定）即失去主要支撑。

7.4 三项研究共同的方法学要点

以下五点适用于三项研究，写在此处以免重复。

其一，盲评。 所有主要终点由不知分组、不知假设的评分者独立评定，报告评分者间信度（ICC，双向随机、绝对一致）。评分材料随机化呈现顺序并去除任何可推知条件的线索（如场地标线、节拍器声）。

其二，材料平衡。 每项研究准备四套等难度序列，按拉丁方在条件间轮换，并在预实验中以独立样本核验四套的等难度性（学习曲线无显著差异）。**这一步不可省：**舞蹈序列的难度差异极易掩盖中等效应。

其三，操纵检查前置。 负荷等同（研究二）与可失败性差异（全部研究）必须作为**独立的检查**先行报告，且预注册中写明：**操纵检查失败时，主分析不作为对假设的检验报告。** 这一条防止把失败的操纵读成阴性结果。

其四，保持间隔固定为 24 小时。 心理练习文献显示效应随间隔衰减且衰减速率不同，因而间隔必须固定并预注册；若资源允许，另加 7 天延迟测以检验衰减速率的组间差异——本文预测**去约束条件的衰减更快**，这是一条额外的、免费的检验。

其五，全部数据与分析脚本公开。 鉴于本文的主张之一是“想象里的东西不会失败”，把数据与脚本置于可被他人推翻的位置，是本文自身唯一诚实的做法。

7.5 三项研究的次序建议

若只能做一项，做**研究三**（最便宜、最尖锐）。若能做两项，加**研究二**（唯一的分界点）。研究一最贵、也最容易被“任务特异”的辩护消解，宜最后做或交由多中心重复。

八、与最近的既有解释的分界

本节的要求是：对每一个可能吸收本文的既有构念，给出一个**两者会作出相反预测的具体情形**。仅指出“侧重不同”不算划界。

8.1 刻意练习的即时反馈要求（Ericsson 一路）

刻意练习的定义中包含“即时反馈”与“超出当前水平的任务”，这与本文关切高度重叠。（敌意检索之后，本文认为最危险的一位是 8.6 的保真度之争，而非本节。）

分界：刻意练习中的反馈是**被给予的**——由教练、教师或明确的成绩指标提供；而本文所说的失败是**被撞上的**——由练习结构中的外部约束自动产生，不需要任何人告知。二者在教学条件下常常同时出现，因而难以分开。

判决性对照：在**完全没有教练、没有外部评价、没有成绩记录**的自主练习条件下，操纵可失败性。刻意练习框架预测：缺少即时反馈时，练习质量普遍低下且不因约束条件而异；本文预测：即使无人反馈，保留外部约束者仍显著优于未保留者。研究二的乙组正是“无教练无反馈”的条件，其与甲组的差异即此对照。

进一步的分歧：刻意练习要求任务**超出当前水平**（难度递增）；本文不要求难度递增，只要求**失败可能非零**。一个远低于当前水平但仍可能出错的练习（如熟练舞者与新同伴合队形），按刻意练习框架价值有限，按本文仍有价值。这是可以直接比较的。

8.2 合意困难（Bjork 一路）

合意困难主张某些增加练习期间困难的操作（间隔、交错、提取而非重读）损害即时表现而改善长期保持。

分界一（自变量层级）：合意困难的自变量是**练习的调度**（何时练、按什么顺序练、练与测的间隔），本文的自变量是**单次练习内保留了什么**。二者正交，可同时施加。

分界二（形状）：合意困难族的典型预测是单调的——在可承受范围内越难越好；本文预测倒 U，且拐点由"失败是否可能"而非"难度高低"决定。

判决性对照：构造一个**更难但不可失败**的条件（如：要求以极慢速度想象整段并保持极高清晰度——认知上更难，但仍然不会出错）。合意困难预测其优于容易的想象；本文预测其不优于，甚至因资源被占用而更差。

8.3 误差驱动的运动学习与内部模型（Wolpert 一路）

该框架主张运动学习由感觉预测误差驱动，前向模型据误差更新。

分界：那一族把误差当作**学习的输入信号**并研究其如何被处理；本文把可失败性当作**练习设计的自变量**并追问它何时被结构性地清零。二者不冲突——本文可以被读作对该框架的一个设计层推论。但它们在一处分歧：

判决性对照：想象练习中，若被试报告了清晰的"想象出的误差"（例如想象自己在第 17 拍踩空），误差驱动框架预测这足以驱动学习（因为误差信号存在于内部模型层）；本文预测**不足**——因为该误差由被试自己生成，可被同时重设，不构成对约束的违反。可通过对比"自生想象误差"与"外源想象误差"（研究三组2、组3）来检验：本文预测后者显著优于前者。

8.4 功能等价假说

已在 2.3 处理。核心分歧一句话：**它把想象的不足归为"弱"，本文归为"不可失败"**。判决性对照即研究三——若不足是"弱"，则提高生动度应改善之，而植入外部约束不应改善之；本文预测相反。

8.5 无误学习（errorless learning）

无误学习主张在某些人群（记忆障碍者、部分技能初学者）中，避免错误的练习优于试错。这似乎与本文相反。

分界：无误学习的适用条件是**错误会被作为内容记住**（内隐记忆保留了错误反应本身），因而其对象是**内容层的错误**；本文所说的失败是**约束违反**，其信息价值在于标定边界而非提供内容。

判决性对照：在无误学习有效的人群上，比较"内容无误但仍会撞到约束"与"内容无误且不会撞到任何约束"两种练习。无误学习框架预测两者相当；本文预测前者仍优。

同时这构成对本文的一条限制：在错误会被记住为内容的情形下，本文的处方须让位。见第十节。

8.6 模拟训练的保真度之争（本文最近的一位，此前被本文遗漏）

这是敌意检索后补入的一节，也是本文必须最用力对付的一位。

模拟训练研究早已发现"更像不一定更好"。该领域把保真度拆成**物理保真**（看起来像）与**功能保真**（行为像），并有影响很大的主张认为应当放弃"保真度"这个词、改用"物理相似性"与**功能任务对齐**，因为物理相似性与迁移的关系微弱。更尖锐的是若干随机对照结果：在急救训练中高保真组的表现并不优于低保真组，某些子项上低保真组反而更好，且高保真组出现了**过度自信**。

必须承认的重叠：本文关于"削减不一定有害、关键在于保留了什么"的那一半，在这一支里已有一个成熟答案——保留的应当是**功能对齐**。本文的"执行量／保留内容"之分，与"物理保真／功能保真"之分在结构上同构。本文此前未提及这一支，是一处实质性疏漏。

分界：功能保真仍然是一个**相似度**概念，只是把相似的维度从外观换成功能；就类型而言它属于特异性原则那一族的精细版。本文的自变量不是相似度的任何一种——**可失败余量不要求练习与目标任务在任何维度上相似**。

判决性对照之一（与 8.10 共用同一操纵）：加入一个与目标动作在功能上完全无关的外部约束（如与本次动作节律无关的节拍同步要求）。功能保真说预测这降低功能对齐、因而无益或有害；**本文预测有益**。

判决性对照之二（本文独有）：构造一个**物理与功能保真都很高、但没有任何失败判定**的模拟——外观与响应都逼真，允许无限次撤销、无计时、无判分。功能保真说预测其有效（对齐度高）；**本文预测其落在格IV，效力存在与投入无关的天花板**。高保真组"表现不更好却更自信"的既有结果与本文预测方向一致，但那些研究并未测量失败判定的有无，因此只是提示这个对照值得做，不构成支持。

8.7 想象的不足是"弱"还是别的（表征重构一路）

功能等价假说之外，另有一条路径主张想象与外显执行**并不共享全部神经底物**，想象练习的效力来自**动作表征结构的重组**，而非来自弱化的执行。

这一支在否定项上与本文一致——都不认为想象只是"强度低的执行"——因而是**同盟兼竞争者**。**分界在肯定项：**那一支说想象靠重组表征起作用；本文说想象的上限由不可失败决定。二者可同时为真，但预测不同。

判决性对照：在想象练习中植入外部约束（研究三组2、组3）。表征重构说预测无改善甚至变差（外部干扰打断重组）；**本文预测改善**。这与 8.4 对功能等价假说的对照是同一个实验而竞争对手不同——研究三因此一次可分开三家。

8.8 运动探索与连续失败（本文此前未引的一族）

运动学习中另有一条经验线索：在二元成败反馈的任务中，连续失败并非只是要避免的事件，而是试错过程的组成部分，其出现频率与学习速度相关；另有计算模型主张失败会触发风险规避、把学习切换到"安全模式"。

分界：那一族关心失败发生之后系统怎么反应（探索幅度如何调整），自变量是失败的发生率；本文关心失败在结构上是否可能，自变量是可失败性的有无。二者的差别在 4.5 的测法一与测法二之间已写明：一次练习可以余量为正而恰好一次没错。

判决性对照：构造两个失败发生率相同、而失败来源不同的条件——甲组的失败来自外部约束（越线、错拍），乙组的失败来自任务本身的随机判定（同比例试次被判失败，但判定与执行者的动作无关）。探索一族预测两者相当（失败率相同）；本文预测甲组显著优，因为乙组的失败不指示任何可被修正的偏差。

8.9 生产性失败（Kapur）

生产性失败主张让学习者先无指导下尝试并失败，再给予讲授，其长期效果优于先讲后练。**须先承认其证据分量：**该路径的元分析涵盖数十项研究，作者报告的效应量按其自设基准约为"一个好老师一年"的两倍。本文此前对它的处理过轻。

分界：生产性失败的单位是一段教学序列（先失败、后讲解），且失败是任务级的（问题解不出来）；本文的单位是单次练习内的一个瞬间，且失败是执行级的（错拍、越线）。二者可同时成立且互不覆盖。

判决性对照：生产性失败要求失败之后必须有讲解（否则效果消失）；本文不要求任何讲解。在无后续讲解的条件下，生产性失败预测效应消失，本文预测效应保留。

8.10 约束主导法与生态动力学

该取向主张技能在任务、个体与环境三类约束的交互中自组织形成，实践上主张操纵约束而非规定动作。这与本文的处方外观相近——都指向"操纵约束"。

分界：该取向操纵约束是为了引导出目标动作模式（约束塑形）；本文操纵约束是为了保住出错的可能（约束作为可失败性的载体）。二者在一处分歧：该取向倾向于选择那些能使正确动作更容易出现的约束；本文主张某些不塑形任何动作、只制造出错可能的约束（如与本次动作无关的外部节拍）同样有效。

判决性对照：加入一个与目标动作模式无关、仅提供同步要求的外部节拍。约束塑形取向预测无益或有害（引入无关约束）；本文预测有益。

8.11 引导假说与专长逆转效应

引导假说主张：练习中过多的外部引导（持续反馈、物理辅助、频繁提示）会使学习者依赖之，从而损害撤除后的表现。专长逆转效应主张：对新手有益的支持性设计对专家反而有害。两者都在"辅助有代价"这一点上与本文一致，是心理学侧最容易吸收本文的两位。

分界：两者的自变量都是**外部提供的信息量**——引导多少、支持多少；本文的自变量是**失败的可能性**，而它可以在信息量不变的情况下被单独操纵。

判决性对照：构造一个**信息量高但不可失败**的条件与一个**信息量低但仍可失败**的条件。例如：甲组在标记时获得持续的语音提示（信息量高），同时保留同伴队形约束（可失败）；乙组无任何提示（信息量低），单独练习无约束（不可失败）。引导假说预测乙优于甲（引导越少越好）；本文预测甲优于乙。**这是本文与引导假说方向完全相反的一处，也是最便宜的一处对照。**

一处必要的承认：若该对照跑出乙优于甲，则信息量而非可失败性是主变量，本文应退为引导假说的一个特例。这一条已并入第十节的作废条件之列。

8.12 与同栏《再生核》（之三十三）

同栏同日另有一篇取同样三个学科的文章，必须先划清。那一篇问的是**技能由什么构成**：拿掉哪一样，下一个身体就长不出来——判据是**留一法教学**，即把一项内容从教学里系统性移除，看下一个还长不出来；第二根轴是可不可陈述。本文问的是一次练习被削减到哪一步就不再出得了错——判据是可失败余量，第二根轴是失败不可见。

两篇表面上都在做"拿掉一样东西看会怎样"，但拿掉的对象、时间尺度与失败的判定者完全不同：那一篇拿掉的是**教学内容**，尺度是**跨人、跨代**，判定者是**下一个学习者长不长得出来**；本文拿掉的是**单次执行的一部分**，尺度是**一次练习之内**，判定者是**执行者本人当场**。

判决性对照：取一项**移除后下一个人仍能长出来的内容**（按那一篇不在核里，可以不教），而它在练习中恰好承担着唯一的外部约束（例如与同伴的队形要求——就技能构成而言可有可无，但拿掉它这次练习就不会出错了）。那一篇预测移除无损；本文预测移除会显著降低本次练习的学习收益。反过来，一项处在核里、却不提供任何当场失败可能的内容（例如只在事后讲评中出现的原则），那一篇预测移除有害，本文预测对单次练习无影响。**这两格是两篇可以互相证伪的地方，也说明二者的自变量正交而非同一件事的两种说法。**

8.13 与本组前作的关系

本文与作者前作《静止点》《类内距》在自变量上正交：那两篇分别处理"替换需要什么时刻"与"同类之间有多近"，本文处理"一次练习保留了什么"。三者可在同一系统中各自为真。

九、跨领域兑现

一条判断若只在其出生的三个领域内成立，它仍是那三个领域的内部整理。以下十处按"可测的可失败性读数"给出。

1. **器乐练习。**慢练之所以有效，不只因为它降低难度，而因为它保留了与节拍器、与谱面的咬合。**读数：**无节拍器的慢练（可失败性低）与有节拍器的慢练（高）之间的学习差异。预测：后者显著优，且差异不随速度变化而消失。
2. **外科技能训练。**模拟器的价值取决于其是否会当场判失败（组织撕裂、器械相撞、超时）。仅供"熟悉步骤"而无判定的模拟落在格IV。**读数：**模拟器每小时触发失败判定的次数，与迁移到真实操作的相关。**这一处有现成的自然实验：**同一款模拟器通常可切换"练习模式"（无判定、可随时重来）与"考核模式"（有判定、计分）。本文预测在总时长相等时，考核模式下的练习迁移更好；而"降低压力有利于学习"的通行主张预测相反。医院技能中心的既有日志足以做这个回溯比较。
3. **飞行训练。**程序演练（chair flying）之所以被要求配以口令与计时，正是把 F 从零抬起。**读数：**有无外部计时对程序熟练度迁移的影响。
4. **语言学习。**默读与影子跟读（shadowing）的差别不在强度而在可失败性——跟读会跟丢。**读数：**跟丢次数与后测口语流利度的相关；本文预测正相关（在中等区间）。
5. **公开演讲准备。**在脑中过一遍（格IV）与对着计时器与真人过一遍（格II）的差别。**读数：**准备过程中"卡住"的次数。
6. **代码学习。**读教程与照抄不同，照抄与自己写不同——差别在编译器会不会报错。**读数：**练习过程中的错误编译次数；预测其与迁移测验成绩在中等区间内正相关，高区间转负（进入格I的资源挤占）。**这一处在当下有额外的意义：**当补全工具能在错误发生前就把它抹掉时，练习被系统性地推向格IV——代码仍然在被写出来，而"写错"这件事不再发生。按本文，这是**执行量不变而可失败余量归零**的一个纯样本。可检验的读数是：在补全强度不同的两组学习者之间，比较**脱离工具后**的迁移成绩，而不是比较他们带工具时的产出速度。
7. **驾驶训练。**空场地绕桩（格II：桩会被撞倒）与模拟软件里无并置判定的路线演练（格IV）。
8. **急救与消防演练。**有计时、有随机变量注入的演练与"按流程走一遍"的演练。**读数：**演练中出现意外偏离的次数。
9. **军事沙盘与桌面推演。**有红队（会反制）与无红队（自说自话）的推演；后者在结构上不会失败。**读数：**推演中方案被迫修改的次数；本文预测该次数而非推演时长预测后续实兵演习表现。这一处与格IV的关系尤其清楚：一场没有红队的推演，其结论在结束时**必然是**"我方方案可行"——因为没有任何东西能否定它。
10. **康复训练——反向约束（第一条让步）。**见第十节。

9.1 两处反过来给本文加约束

第一条：康复与高风险人群。在跌倒风险高的康复训练中，制造可失败性会直接造成伤害；同时（见8.5）在部分记忆障碍人群中错误会被记为内容。因此本文的处方在这两类人群上**必须让位**，且4.2的三个构成条件（失败代价足够低）在此不满足。

这条约束反过来强化了框架：它说明可失败性不是一个越高越好的量，而是一个**必须先有安全垫才能被使用的量**。由此得到一条本文原本没有的推论：**提高安全性是提高可学习性的前置投资，而不是与之权衡的对立项**。护具、软垫、可撤销的操作、沙箱环境——它们的价值不只是防止伤害，而是把可失败余量从"不敢用"变成"可以用"。

第二条：表演型任务的最终阶段。 在临近演出、比赛或手术的最后阶段，练习的目的从"学会"转为"确认与稳定"，此时特异性原则重新占据主导——需要在与目标条件尽可能一致的条件下完成完整执行。因此本文的适用范围是**技能获得阶段**，而非**赛前调整阶段**。

这条约束也反过来给出了一个可检验的预测：在同一支舞队的排练周期内，标记相对于全力排练的优势应当随演出临近而单调下降，并在最后一周内消失或反转。这是一个用现有排练日志即可回溯检验的预测，且没有任何现有理论作此预测。

十、适用边界、反噬预言与使本文为假的条件

10.1 适用边界

其一，**失败代价必须低**（4.2 第三条）。高风险与易受伤情形不适用。

其二，**错误不得被记为内容**（8.5）。在内隐记忆保留错误反应本身的人群与任务上不适用。

其三，**技能获得阶段**，非赛前调整阶段（9.1 第二条）。

其四，**观摩、阅读与讲解不在本文覆盖之内**（5.2 反例三）。它们经由另一条通路起作用，本文对其效力上限不作主张。

其五，**执行者须具备识别失败的能力**。这一条对新手是实质限制：新手可能撞到约束而不自知。因此本文预测可失败性的效应**随专长上升而增强**——这也是 H5 的一个推论，并与心理练习的专长调节效应方向一致。

10.2 反噬预言

按本文设计干预，最自然的做法是要求练习计划中报告"每次训练的失败次数"。

这会适得其反，且方式可以精确说出：失败次数是一个**极易被制造**的量。最省事的达标办法不是保住有意义的约束，而是**引入任意的、与技能无关的约束**（无关的计时、随意的边界线、人为的干扰音），使失败次数上升而学习不变。更糟的是，一旦失败次数成为绩效指标，执行者会转而选择那些**失败容易发生但容易被消除**的约束，从而把练习推向"制造并消除人为障碍"的空转。

因此本文不建议把失败次数做成申报指标，而建议一件不能被优化的事：**随机抽一次练习，问执行者"刚才哪一处你知道自己错了"**，并核对该处是否与目标技能相关。这个动作测的是失败的**内容**而非数量，而内容无法被凑数。

10.3 两条使本文为假的条件

其一（关键实验）：在负荷成功等同（TOST 通过）的前提下，保约束组与去约束组在学习收益上无显著差异且等价性检验支持等价（研究二）。此时本文应被并入具身认知负荷假说。

其二（想象植入）：在纯想象练习中植入外部约束（外部节拍、随机中断）后，学习效应未高于常规想象（研究三）。此时本文的核心机制失去主要支撑。

其三（方向相反的对照）：在 8.8 所述的"高信息量+可失败"与"低信息量+不可失败"对照中，若后者优于前者，则信息量是主变量，本文应退为引导假说的特例。

三条相互独立：第一条针对"资源之外还有别的东西"，第二条针对"那个东西是可失败性"，第三条针对"它是否只是信息量的另一种说法"。

10.4 两个交出去的洞

第一，"失败"的边界由谁定。与外部节拍差 30 毫秒算不算失败？本文以"执行者当场察觉"为判据，但察觉阈本身随专长与注意状态变化，因而可失败余量在跨人比较时不可通约。本文没有独立的定阈办法。

第二，本文未处理"失败的类型"。撞到同伴、错拍、失衡在本文中被同等对待，而它们很可能对不同的学习成分起作用。这个洞是实质性的：若不同类型的失败作用于不同层，则本文的单一构念应当被拆开。本文不填这个洞，并把它列为最可能的后续修正方向。

十一、实践含义

含义一：把"练得像"换成"还能不能错"。训练设计的第一问不应是"这个练习像不像比赛"，而应是"这个练习里，人还会不会当场知道自己做错了"。前者常常导向昂贵而疲劳的全力重复，后者导向便宜而高效的保约束标记。

含义二：独自练习的结构性缺陷。一个人在没有任何外部约束的条件下练习，无论多么投入，都很容易落进格Ⅲ或格Ⅳ。最便宜的补救不是增加时长或强度，而是**加一个不由自己控制的东西**：节拍器、计时、录像回看的固定检查点、一个同伴。

含义三：视频与讲解的天花板。观看示范与阅读讲解在结构上不会失败，因而其效力有一个不随投入增加而抬高的上限。本文不主张它们无效，而主张**任何以它们为主体的学习安排都停在格Ⅳ**，且这一点无法通过"看得更认真"来克服。

含义四：安全是学习的前置条件而非其代价。见 9.1 第一条。

含义五：辅助工具的两面性。任何降低出错概率的辅助——护栏、自动纠正、补全、提示、导航——在提高当次表现的同时降低可失败余量。本文不主张取消它们，而主张**把"带辅助的表现"与"撤辅助后的能力"分开测量**，并把后者作为学习是否发生的判据。一个只在带辅助时被测量的体系，在结构上无法发现自己已经进入格Ⅳ。

11.1 一条关于剂量的推论

本文给出一条与通行做法相反的剂量建议。通行做法把"全力完成"作为默认，把削减作为疲劳或伤病时的退让。按本文，默认应当反过来：

默认采用刚好还能出错的最省力版本；把全力执行留给两个特定用途——一是检验（看看在真实负荷下是否仍成立），二是赛前的稳定期（9.1 第二条）。

这条建议成本极低而后果不小：它把大量高负荷、高疲劳、低学习的重复（格 I 的低效部分与格 III 的全部）替换为低负荷、保约束的练习，同时**不减少甚至增加当场失败的次数**——因为失败来自约束而不是来自负荷。

它在两个条件下不成立：体能适应本身是目的时（5.2 反例一），以及临近目标事件时（9.1 第二条）。

十二、本判断对它自己适用吗

本文自己是一次练习吗？

按本文的判据，一篇理论论文的写作落在**格IV**：它在原理上不会当场失败。论证中的每一处困难都由作者自己设定，也由作者自己解除；没有一个不由作者控制的约束会在写作过程中当场判定"这里错了"。这正是本文所指认的那一格。

本文对此的唯一补救，是把可失败性写进结构：第七节三项研究的**预注册证伪条件**、第十节两条使本文为假的条件，都是**由外部数据而非由作者判定**的失败通道。它们的作用不是让论证更有说服力，而是在**本文之外造出一个能判它错的地方**。

但必须诚实地说：**在有人真的执行研究二或研究三之前，这些通道只是被写下来的，不是被打开的。** 一条从未被使用过的失败通道与不存在的失败通道，在结构上无法区分——这正是本文第五节对格IV所说的话，此刻适用于本文自身。因此本文的正确地位是：**一个提出了可失败结构、但尚未被失败过的理论**。

第二重自反更具体。本文的核心证据来自两项独立文献的并置，而并置本身是我在写作中完成的一次"想象"——两项结果从未在同一实验中被观察到。按本文自己的框架，这次并置属于自生的推论而非撞上的约束。**研究一存在的全部理由，就是把这次并置从想象里搬出来。**

十三、投稿路线与本稿现有的不足

13.1 目标期刊

首选：Human Movement Science (Elsevier, SSCI Q2)。理由：该刊接受理论/概念性论文，且长期发表运动学习、练习设计与舞蹈科学交叉的稿件；本文的三学科结构与其读者匹配。

备选一：Research Quarterly for Exercise and Sport (SSCI Q2)。理由：接受"Viewpoint / Theoretical"栏目，篇幅与本文相近。

备选二：Frontiers in Psychology (Movement Science and Sport Psychology 专栏, SSCI Q2)。理由：接受 Hypothesis and Theory 类型，明确容纳"提出可检验假设但尚无数据"的稿件——**就本稿当前形态而言，这是命中率最高的一个。**

若先完成研究三（成本最低），可上冲 Psychology of Sport and Exercise (Q1) 或 Journal of Experimental Psychology: Human Perception and Performance (Q1)。

不建议：Journal of Dance Medicine & Science（分区偏低，且会把本文框成舞蹈专业问题，损失其跨领域主张）。

13.2 本稿距离可投出还缺什么

诚实列出，不掩饰：

1. **没有数据。** 纯理论稿在 Q2 的接受率显著低于实证稿。最低可行方案是先做**研究三**（ $N \approx 180$ ，可在体育院校一学期内完成，无需舞蹈场地），把本文改写为"理论 + 一项实验"。

2. **两处文献需回到原文核对。** 特异性原则的表述依据 Proteau 一系列工作的公认结论，本稿**未逐条核对原文页码与具体实验参数**；Warburton 等人 2013 的材料细节（动作元素数量、评分方式）本稿依据公开摘要与二手评述，投稿前必须取原文核对并按其实际参数重写 7.1 的材料部分。

3. **英文改写不是翻译。** 目标刊的 Introduction 通常要求在前两段内给出 gap 与贡献；本文第一节的中文写法偏叙述，英文版需重组为"三条结论 → 不一致 → 本文的构念 → 关键实验"四段式，并把第三节的表格前移到 Introduction 末尾。

4. **伦理与预注册。** 三项研究均需伦理审批号；建议在 OSF 预注册后再投，并在稿中给出注册号——对本文这类"以证伪条件为卖点"的稿件，预注册是可信度的主要来源。

5. **样本量估算依赖一个借来的效应量。** 研究二的 $d = 0.6$ 取自 Warburton 等人组间差异的量级，而该项研究的操纵（标记 vs 全力）与本研究的操纵（保约束 vs 去约束）并不相同，**这个借用是本稿最脆弱的方法学环节**。投稿前应是一项小规模预实验（每组 $n \approx 15$ ）重估效应量，或按更保守的 $d = 0.45$ 重算（每组 $n \approx 79$ ）。

6. **保真度那一支是补入的，尚未消化完全。** 第 8.6 节是在敌意检索后补写的；该文献量极大（医学教育、航空、军事各有独立传统），本稿只处理了其主干主张。投稿前应至少补读一篇系统综述，并考虑把第九节第 2 条（外科模拟器）改写为直接对话该文献的形式。

7. **"可失败余量"的英文命名需要斟酌。** failability 一词在英文中不常用，审稿人可能建议改为 error-affordance 或 failure availability。本文倾向保留 failability margin 并在首次出现处给出定义，但这一点应做好在修回中让步的准备。

13.3 本文自认的位置，以及为什么这里不写一个数

本文在设计上做齐了四件事：与既有说法逐一划界并各配一个方向相反的判决性对照（第八节，十二位）；跨领域兑现十处，其中两处反过来给本文加约束（第九节）；自反一节真认账（第十二节）；以及一条写死日期的赌注：

到 2031 年 12 月 31 日，若研究三（想象中植入外部约束）已被至少一个独立团队执行，且植入组未显著优于常规想象组，则本文的核心机制作废，不再申辩。

本文不在此处写任何评分。理由很简单，也与本文的主张同构：一个由作者自己给自己的判定，在结构上不会失败——它落在第五节的格IV里。要有一个算数的数，必须由不知道本文推导过程的人独立给出，或者由第七节那三项研究给出。在那之前，本文所能诚实主张的只是“它给出了一条可检验的区分”，而不是“它有多好”。

参考文献

- Bjork, R. A. (1994). Memory and metamemory considerations in the training of human beings. In J. Metcalfe & A. Shimamura (Eds.), *Metacognition: Knowing about Knowing* (pp. 185–205). MIT Press.
- 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.
- Feltz, D. L., & Landers, D. M. (1983). The effects of mental practice on motor skill learning and performance: A meta-analysis. *Journal of Sport Psychology*, 5(1), 25–57.
- Hamstra, S. J., Brydges, R., Hatala, R., Zendejas, B., & Cook, D. A. (2014). Reconsidering fidelity in simulation-based training. *Academic Medicine*, 89(3), 387–392.
- Kapur, M. (2008). Productive failure. *Cognition and Instruction*, 26(3), 379–424.
- Massoth, C., Röder, H., Ohlenburg, H., Hessler, M., Zarbock, A., Pöpping, D. M., & Wenk, M. (2019). High-fidelity is not superior to low-fidelity simulation but leads to overconfidence in medical students. *BMC Medical Education*, 19, 29.
- Norman, G., Dore, K., & Grierson, L. (2012). The minimal relationship between simulation fidelity and transfer of learning. *Medical Education*, 46(7), 636–647.
- Newell, K. M. (1986). Constraints on the development of coordination. In M. G. Wade & H. T. A. Whiting (Eds.), *Motor Development in Children: Aspects of Coordination and Control* (pp. 341–360). Martinus Nijhoff.

- 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.
- Proteau, L., Marteniuk, R. G., & Lévesque, L. (1992). A sensorimotor basis for motor learning: Evidence indicating specificity of practice. *The Quarterly Journal of Experimental Psychology A*, 44(3), 557–575.
- Toth, A. J., McNeill, E., Hayes, K., Moran, A. P., & Campbell, M. (2020). Does mental practice still enhance performance? A 24-year follow-up and meta-analytic replication and extension. *Psychology of Sport and Exercise*, 48, 101672.
- 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.
- Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12), 739–751.
- Wulf, G., & Shea, C. H. (2002). Principles derived from the study of simple skills do not generalize to complex skill learning. *Psychonomic Bulletin & Review*, 9(2), 185–211.

人机分工声明

本文由王德生与 Claude 共同署名。三个学科由王德生指定（体育+舞蹈艺术+心理学）；三家对立理论的确立、冲突的形式化、构念的提出、判别表、六条假设、三项研究方案（含样本量估算与预注册证伪条件）、七处理论分界与十处跨领域兑现，以及全部文字，由 Claude 生成；方法论框架与发表裁定由王德生给出。

核验状态：Warburton et al. (2013) 与 Driskell et al. (1994) 的结论与关键数值经检索核对；Proteau 一路的表述依据其公认核心结论，**未逐条核对原文页码与实验参数**；其余条目引用限于公认论点。**本文未收集任何数据，第七节三项研究给出的是方案而非结果**，这一限制已写入第十二节与第十三节。

字数 纯汉字约 1.9 万 **版本** v1.0 · 2026-08-01

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