

交出去的不止那件事

为什么每一次外包，都顺手外包掉了你验收它的能力

王德生 + Claude · 约 2.0 万字 · 18 页 · 三种阅读方式 · 发表于2026年7月28日

一位老师把「这学生长得怎么样」交给了分数；一个团体把「这信是不是真的」交给了可以数出来的东西；一个人把「我身体怎么样」交给了手环。三次交出去，理由都充分。本文认为每一次交出去的都不止那一样：判断这件事做得好不不好的能力也跟着走了——因为那个能力本来就是「自己做」的副产品。于是每一次委托都同时委托掉了对这次委托的监督权，而这是唯一一种你不会发现的损失：一旦发生，用来发现它的那个器官已经不在了。

三件被交出去的事

先看三件事。它们分属三个领域，彼此没有关系。

第一件，把「这个学生长得怎么样」交给了分数。

一位老师心里清楚班上某个学生这半年发生了什么变化——他开始会自己回头检查了，他遇到不会的题不再第一时间抬头看老师了，他有一次为了一个想法争了整节课。这些东西没有一样能填进成绩单。于是最后被送出去、被用来决定这个学生去哪儿的，是一串分数。

半个世纪以来，人们一直在改进这串分数：加等级、加评语、加过程性评价、加成长档案。每一次改进都更精确一点，每一次都失败了。而失败之后得出的结论永远是同一个：**代理品还不够好，我们需要一个更好的代理品。**

第二件，把「这信是不是真的」交给了可以数出来的东西。

一个宗教团体面对一个极难回答的问题：我们怎么知道这些是真的？这个问题在内部无法用任何可观测的方式回答——它涉及的东西不可见、不可控、不可量化。

于是慢慢地，一套可以观测的东西被建立起来：一套人人会讲的个人经历叙事，一套关于「感受到了什么」的判断方式，一套可以统计的数字——聚会人数、奉献金额、参与率、事工完成度。这些东西替那个无法回答的问题，提供了一个可以拿在手里的答案。

要紧的是：这不是骗局。它回答的是一个真实的、迫切的、不回答就撑不下去的不安。

第三件，把「我身体怎么样」交给了手环和平台。

一个人以前判断自己的状态，靠的是一套模糊但一直在用的感觉：昨晚睡得沉不沉、爬楼梯到第几层开始喘、下午几点开始发飘。现在这些都有了读数：深睡几分钟、静息心率多少、步数达标没有。

于是那套模糊的感觉慢慢就不用了。不是他被禁止用，是他不需要用了——有更准的东西在。

而这里最难受的一点：**他知道**。大量的人非常清楚地知道，长期依赖这些读数会让自己那套感觉退化。他们说得出道理，甚至能举出例子。然后他们继续。

三件事，三个领域，一个共同点：**有一样东西被交了出去，而交出去的理由每一次都是充分的。**

这篇文章要说的是：每一次交出去的，都不止那一样东西。还有一样跟着走了，而且是悄悄走的。

跟着走的那一样东西

先把它说出来，再论证。

你把一件事交出去的时候，同时交出去的，是你判断这件事做得好不好的能力。

这句话乍看像老生常谈——不就是「用进废退」吗？不完全是。用进废退说的是那件事本身你不做就生疏了。这里说的是另一样东西：**不是做的能力，是验的能力。**

这两样不一样。

做饭的能力和尝出咸淡的能力不一样。一个人可以几十年不下厨，仍然一口就吃出这道菜火候过了——因为他年轻时做过很多年。反过来，一个从来没做过饭的人，即使天天吃，也说不出这道菜到底哪里不对，他只能说好吃或不好吃。

带过学生的老师，即使自己多年不上课，看一节课仍然知道哪一分钟是活的；从没带过学生的人，看同一节课，只能看出学生举手多不多。

自己写过代码的人，能看出一段代码哪里将来会出事；没写过的人，只能看它跑不跑得起来。

验的能力和做的能力是两件事，但它们有一个关系：验的能力是做出来的。

这里还要防一个混淆：验的能力不是「品味」，也不是「眼光」。这两个词太软，而且暗示某种天赋。

它其实相当具体，具体到可以描述：**它是一张关于「这类事情通常在哪儿出问题」的地图，而这张地图上的每一个标记，都对应着一次你自己踩过的坑。**

一个做过工程的人看图纸，眼睛会不由自主地落在几个地方——不是因为规范要求看那儿，是因为他在那几个地方吃过亏。一个带过团队的人看一份计划，会立刻问一句「这一步谁负责」，因为他有过一次没人负责的经历。

所以这张地图有两个性质，都很要紧。

第一，**它是稀疏的**。它只覆盖你踩过的那些坑，没踩过的地方是空白。所以一个人的判断力永远是有边界的，而且边界的形状取决于他的经历，不取决于他的聪明。

第二，**它不能被转述**。你可以告诉别人「要注意第三步」，但对方接收到的是一条规则，不是一个标记。规则会被遵守，直到出现一个规则没覆盖的情况；标记会让人在那一带整体地警觉起来。这就是为什么经验丰富的人交接工作时总觉得说不完——他要转述的是一张地图，而他手上只有一支笔。

这就是全部问题的根子。你没法只保留验的能力而把做的部分交出去——因为验的能力不是独立长出来的，它是做的副产品。你交出做，验就在若干年后跟着一起走，只是走得慢一点，所以你不会把这两件事联系起来。

为什么验收能力是「做」的副产品

这一节要把上面那句话坐实，否则它只是一个好听的说法。

先问：一个人凭什么能看出一件事做得好不好？

不是靠标准。标准是可以背下来的，而且背下来之后你会发现，符合全部标准的东西照样可能是坏的，违反某条标准的东西照样可能是好的。真正的判断力从来不停在标准上。

它靠的是**你自己做过、并且做砸过**。

具体说，是这样一个过程：你做了一件事，你以为做对了；后来出了问题；你回头找，发现问题出在当初那个你根本没在意的地方。这一次之后，那个地方在你眼里就不一样了——你以后看别人做同样的事，眼睛会自动落到那儿。

这个过程有三个不可省略的环节：

第一，你得亲自做那个决定。不是执行别人的决定，是在一个信息不全的时刻，自己拍板。因为将来出问题时，你要能回想起「我当时是怎么想的」——这个回想是判断力的原材料。执行别人的决定，出了问题你回想起来的是「他让我这么干的」，这不产生任何判断力。

第二，你得承担结果，而且这个结果要真的落到你身上。如果做砸了没有任何后果，或者后果由别人承担，那么你不会去回头找原因。人只有在被咬过之后才会记住哪里有牙。

第三，你得自己发现哪里错了，而不是被告知。这一条最容易被忽略。被告知「你当初那一步错了」，你学到的是一条结论；自己找出来，你学到的是一套找法。以后遇到没人告诉你的情况，结论用不上，找法用得上。

三个环节合起来，才长出一点点判断力。而它们全都要求：**这件事得是你自己做的**。

现在回头看「交出去」这个动作。你把一件事交给别人（或者交给一套系统、一个平台、一串指标），这三个环节就同时断了：决定不是你拍的，结果不直接落在你身上，出了问题也是对方来告诉你哪儿错了。

于是这件事上的判断力，从交出去那一刻起，就停止生长了。已经长出来的那部分还在，会慢慢钝，但不会立刻消失——**所以刚交出去的时候，一切都很好**。你还看得懂，还挑得出毛病，还能把人叫来说这里不对。

问题出在五年后、十年后。

这是唯一一种你不会发现的损失

大多数损失是可以被发现的。钱少了会知道，人走了会知道，东西坏了会知道。发现之后，可以决定修还是不修。

判断力的损失不属于这一类。因为**发现「这件事做坏了」这个动作本身，需要用到的正是那个已经不在的能力**。

这句话值得停一下。

你把评估学生的事交给了一套分数体系。五年后，这套体系开始出偏差——它把某一类学生系统性地评低了。谁来发现？要发现这件事，得有人能在不看分数的情况下，看着那个学生说「他不是这个水平」。而这个能力，恰恰是在这五年里因为不再需要而钝掉的。

你把「信是不是真的」交给了那套可观测的操作。多年之后，那套操作开始与它本来要偿还的东西脱节。谁来发现？要发现，得有人还保有一套不依赖那些数字的判断方式。而那套方式因为多年没被使用过，已经没人再提起了。

你把身体的状况交给了读数。若干年后，读数正常而你确实不对劲。谁来发现？发现需要的正是那套被读数替换掉的模糊感觉。

每一次都是同一个结构：用来发现损失的器官，恰好就是被损失掉的那一个。

所以这种损失不会以「出事了」的形式出现。它出现的形式是：**一切正常，直到很久之后一次谁也没预料到的崩掉**。而崩掉之后的复盘会得出结论：是执行不到位、是标准不够细、是某个环节出了纰漏。因为复盘用的还是那套已经钝掉的判断力，它只能看见它还看得见的东西。

这里还有一层更冷的：**在这种损失里，没有一个时刻是可以被指认为"出错"的**。每一步都是合理的：交出去是为了效率，用指标是为了公平，依赖读数是因为它确实更准。你回头找那个错误的决定，找不到，因为它不存在。

这里还要防一个安慰性的想法：也许可以靠别人来发现？比如让外部的人来看，或者让新来的人来看。

外部的人有一个真实的优势：他没有被这套体系训练过，所以他看得见一些内部人看不见的东西。但他也有一个对称的劣势：他没有在这个现场做过，所以他看不见那些只有做过才看得见的东西。于是他能指出的，往往是那些**在任何地方都成立的一般性问题**——流程冗余、沟通不畅、激励错配。这些诊断通常都对，也通常不解决问题。

新来的人则更麻烦：他会在很短的时间里被同化。不是因为有人压他，而是因为他要在这里做事，就得学会这里的做法，而学会这里的做法就包括接受这里的判断标准。他那点新鲜的疑问，通常在半年内消失，而且他自己会把这个过程记作「我成长了，当初想得太简单」。

能发现这个损失的人，需要同时满足两个条件：在这个现场真的做过，并且没有被这套体系同化。这两个条件在同一个人身上很少同时满足，而且它们随时间朝相反方向走——待得越久越懂现场，也越被同化。

于是「找一个更好的代理」永远打空

现在可以解释开头那件事：为什么半个世纪以来，所有改进代理品的努力都失败了，而人们仍然认为出路是找一个更好的代理品。

因为**改进代理品这件事本身，是由那个已经钝掉的判断力来主持的**。

想一想改革是怎么发生的：有人觉得现在的评价方式不好，于是设计一个更好的。他凭什么知道新的更好？凭他的判断。而他的判断，是在这套已经运行多年的代理体系里长起来的——他看得见这套体系看得见的东西，看不见它看不见的东西。

于是他设计出来的新代理品，会比旧的更精确、更全面、更能捕捉那些**在旧体系里被抱怨过的东西**。它捕捉不到的，恰好是从来没有人抱怨过的部分——因为抱怨也需要先看见。

这解释了一个反复出现的现象：每一次评价改革都确实解决了上一轮被指出的问题，而整体上什么也没变好。**不是改革无效，是改革一直在解决那些已经能被看见的问题**。

还有一个更棘手的版本：改革本身会加深依赖。

一套更精细的评价体系，需要更多的记录、更多的填报、更多的对齐。这些工作会占掉原本可能用来「自己做、自己判、自己承担」的时间。于是每一轮为了修复判断力而做的改革，都在减少判断力赖以生长的那点余地。

这一点必须说清楚，否则会被读成对改革者的指责。改革者做的每一步都对：发现问题、设计方案、更精确、更公平。这条路上没有一个错误的决定。它只是整体地朝着一个方向去，而那个方向没有人选过。

为什么知道了也不会动

到这里要处理一个更难的问题。

上面的分析可能会让人以为：这是一个认知问题——人们不知道交出去的代价，所以才交。那么解决办法就是让他们知道。

但事实是：**很多人知道，而且知道得很清楚。**

大量依赖健康读数的人，能非常准确地说出「我现在已经不太听得懂自己身体了」。他们不是无知。他们知道，然后照做不误。

大量的老师能清楚地说出分数看不见什么，他们甚至能举出具体的学生。然后他们继续按分数排名。

大量的人清楚一个团体正在用可数的东西替换不可数的东西。然后他们继续参加聚会、继续填报表。

所以这不是认知问题。把道理讲得更清楚，不会改变行为。

那是为什么？

我认为关键在于：**要交出去的那个东西，此刻就要一个答案；而不交出去的代价，要很久以后才显出来，而且显出来的时候，你已经无法把它归给这次交出去。**

这两件事不在同一个账户上结算。

「我怎么知道这学生行不行」——这个问题今天必须有答案，因为今天就要填志愿。「我怎么知道这信是不是真的」——这个问题此刻必须有答案，因为不安是此刻的。「我最近状态怎么样」——这个问题今晚就想知道。

而代价是：五年后你不再能自己判断了。五年后的那个你，不会觉得这是当年那个决定造成的——他会觉得这就是他的水平。

一个此刻确定的答案，对一个此刻不安的人来说，几乎是不可抗拒的。

还有一个更硬的原因，值得单独说：**不交出去这个选项，往往在现实中并不存在。**

前面的讨论有一个隐含假设，好像每个人面前都摆着两个选项：交出去，或者不交。实际上大多数时候只有一个。

那位老师不是在「按分数评」和「按自己的判断评」之间选择——升学要用分数，家长要看分数，学校的考核也是分数。她能选的只是要不要在分数之外**额外**花一份心力，而那份心力不进入任何账目。

那个人不是在「用手环」和「不用手环」之间选——他的体检报告、他的医生、他的保险，说的都是同一套读数的语言。他即使扔掉手环，也还是要用那套语言跟所有人交流自己的身体。

所以这里说的「交出去」，很多时候不是一次选择，是一个**已经完成的、由整个环境构成的既成事实**。个人能做的只是在这个事实之上，自己再花一笔额外的钱去保住一点东西。

这一点让前面的分析变得更冷：**这不是一个可以靠个人清醒来解决的问题**。一个人再清醒，他也改不了升学要看分数、体检要看指标、汇报要看报表。他能做的只有第十四节那些代价由他自己承担的小动作。

把这件事说成个人选择，其实是把一个结构问题的账，记到了个人的头上。而这里最要命的是：那个交出去的动作，**当场就兑现了它承诺的东西**。它不是空头支票，它真的给了你确定性。所以它不会被识别为一次损失，它会被识别为一次得到。

为什么最内行的人也逃不掉

有人会说：这个问题对外行成立，对真正的内行不成立——一个干了三十年的人，他的判断力已经长成了，交出去也带不走。

这个说法有一半是对的。已经长成的判断力不会因为几年不用就消失，它会钝，但底子还在。所以一位老手在退居二线之后的头几年，仍然是最好的把关人。

但有两件事会同时发生，而它们合起来足以抵消那个底子。

第一件：他要验的东西，已经不是他当年做过的那件事了。

一位老编辑的判断力，是在纸质出版的年代长出来的——那时候一篇稿子的好坏，与它在纸上被人一行一行读下去的方式紧密相关。今天他要验的稿子，是给一个完全不同的阅读场景写的。他的底子仍然在，但底子对应的那套条件变了，而变化的部分恰恰是他没有亲自做过的部分。

他的判断力不是钝了，是错位了。而错位比钝更麻烦：钝了的人知道自己看不清，错位的人看得很清楚——只是看的是另一件事。

第二件：他会开始相信自己的判断力，超过它实际值得的程度。

这是必然的，因为他没有任何机会得到相反的反馈。他在这个位置上不再自己做，也就不再自己做砸；不做砸，就不会有那种「原来我当初想错了」的时刻。而正是那种时刻在一直校准着一个人对自己判断力的估计。

校准一停，估计就只会往上走——不是因为他自负，而是因为**每一次他的判断被采纳、事情也没出大问题，都会被记作一次验证**。而事情没出大问题的原因可能是执行的人在下面默默补了很多，也可能是那个问题的显现期还没到。

这两件事合起来，产生了一个相当难堪的结论：**在一个人的判断力最被倚重的那些年，恰恰是它最缺乏校准的那些年**。而且组织的结构会让这件事更严重——越往上，越少亲自做，越多做判断，越少得到反馈。

这也解释了一个常见的现象：一个部门的问题，往往是底下的人先看出来的，而他们说不动上面。通常的解释是「上面官僚」或者「屁股决定脑袋」。这些解释有时候是对的，但还有一个更朴素的可能：**上面那个人的判断力，确实已经不覆盖底下这件事了，而他没有办法知道这一点**。

于是「交出去」变成了默认

再往前一步。

如果只是「交出去有诱惑力」，那还只是个体选择的问题。但实际情况比这更硬：在多数场合里，**交出去已经成了默认路径，而不交出去成了需要辩解的例外**。

一位老师如果说「这个学生我不按分数评，我看着他这一年长起来的」——他必须解释。他要说清凭什么、依据是什么、如何保证公平、如果家长质疑怎么办。这些质问都合理，而且他往往答不上来，因为他的判断恰恰是那种说不清依据的判断。

一位医生如果说「他的指标都正常，但我觉得不对，再观察两天」——他要担责任。如果两天后确实没事，他浪费了床位；如果出了事，追责会问他为什么不早处理。

举证责任翻转了。过去是「你凭什么要用这个数字代替我的判断」，现在是「你凭什么不用这个数字」。

这个翻转是靠一个非常有力的说法完成的：**可比较才公平**。

这句话很难反驳，而且它大部分时候是对的。没有可比较的标准，评价就变成看人下菜；没有统一指标，权力就会在含糊处膨胀；最先受害的永远是那些说不上话的人。这不是小事，这是极其昂贵的成果，任何轻率地拿它开刀的主张都是拿别人的处境冒险。

但「可比较才公平」有一个未被言明的推论悄悄搭了便车：**凡是不能被比较的，就不算数**。

前半句是对的，后半句是偷渡进来的。而正是后半句，把「交出去」变成了不需要理由的默认动作。

四种交出去，只有一种是安全的

到这里可以做一个有用的区分。不是所有的交出去都会带走验收能力。差别在两个变量上：**你还做不做，以及你还验不验。**

形态	做	验	状态	会滑向
第一种	交出去	自己验，且方式独立于对方的交付物	安全，分工的正常形态	第二种
第二种	交出去	交给与执行方分开的第三方	较弱，仍可用	第三种
第三种	交出去	交给执行方自评	结构上不成立	第四种
第四种	交出去	能力已无，而自以为还有	最坏，也最难修	—

第一种：验还在自己手上。

你不再亲自做，但你仍然定期用自己的方式检查结果，而且你的检查方式不依赖对方提供的读数。这一种是安全的，而且它是分工的正常形态——一位老木匠不再自己做每一件活，但他上手一摸就知道行不行。

维持这一种有一个硬条件：**你的验收方式必须独立于对方的交付方式。**如果你只能通过看对方给的报表来检查对方，那你验的不是那件事，你验的是报表。

第二种：验交给分开的第三方。

这就是审计、第三方评估、外部检查的形态。它比第一种弱，但仍然可用——只要那个验收方自己保有真实的做的经验。

它的失败模式很清楚：验收方久了也会只看材料。因为验收方的考核也是按量、按时、按合规来的，而深入到现场去看是最不划算的做法。审计一旦标准化，它就变成了对材料的审计。

第三种：验也交给同一方。

对方既执行又自评，你只看对方给你的结论。这一种在结构上就不成立，但它极其常见——因为它最省事，而且在初期效果非常好：对方是内行，对方的自评比你的外行判断准得多。

它准，直到它不准。而它不准的那一天你不会知道。

第四种：验的能力已经没有了，而你以为还有。

这是最坏的一种，也是前面几种的终点。你仍然在检查，仍然会挑毛病，仍然觉得自己在把关——但你挑的毛病全都落在那些还能被你看见的地方（格式、时限、明显的错漏），而真正的问题在你的视野之外。

第四种和第三种的区别是：第三种是自愿放弃了验，第四种是以为自己还在验。后者更难修，因为你不会去修一个你认为完好的东西。

这四种之间不是静止的，它们有一个固定的滑动方向：**第一种会滑向第二种，第二种滑向第三种，第三种滑向第四种，而反向从不自发发生。**

滑动的动力每一步都很朴素。第一种滑向第二种，是因为独立验收太耗时，把它交给一个专职的人更划算。第二种滑向第三种，是因为外部验收方总是不如执行方懂行，久了大家发现听执行方的更靠谱。第三种滑向第四种，只需要时间——当年那批还能验的人退了。

没有哪一步是错的，而且每一步都能省下真实的成本。这就是为什么不需要任何人推动，系统也会自己走到第四种。

要停在前面某一档，必须有人持续地往回推，而往回推的每一步都在增加成本、降低效率、且看不出任何收益。

这个四分法可以直接拿来自问：**在我交出去的这些事情里，哪一件我还能不看对方给的材料就判断出好坏？**如果一件也想不出来，那不是因为对方都很可靠。

一个组织是怎么一层层丢掉的

上面讲的是一个人。放到一个组织里，这件事会以一种更麻烦的方式展开：它是**代际**的。

第一代人是自己做起来的。他们做过，做砸过，自己找出过原因。所以他们既能做，也能验。这时候组织很健康——虽然乱，虽然没什么流程，但每一件事都有人真的懂。

第二代人是被第一代带出来的。他们做，但重要的决定是第一代人拍的；他们也承担结果，但兜底的是上一代。所以他们的判断力长得出来一部分——他们至少在现场，至少见过第一代是怎么判的。这一代通常是组织里最好用的一代：既懂现场，又开始能管事。

真正的断层出现在第三代。

第三代人进来的时候，组织已经有了成熟的流程和指标。他们做的事是被流程定义好的，决定不是他们拍的，出了问题有明确的追责路径和纠正机制。他们非常胜任——各项指标都好看，甚至比第二代当年更规范。

但他们没有机会经历第三节说的那三个环节。于是他们的判断力，是关于**流程的判断力**：这一步合不合规、这份材料齐不齐、这个环节有没有超时。这是一种真的能力，而且组织非常需要它。

它只是不能用来验第一代做的那件事。

而到了第三代开始接管的时候，一个关键的事发生了：**没有人能发现这个差别**。第一代已经退了，第二代也快退了，而第三代不知道自己缺了什么——他们缺的东西从未在任何交接材料里出现过，因为它说不清依据，写不下来。

这个过程有一个非常明显的外部特征，可以拿来自查：**当一个组织开始把「有经验」等同于「在这里待得久」的时候，第三代已经接管了**。因为在第一代人的语言里，经验指的是做过和做砸过；在第三代的语言里，经验指的是熟悉流程。两个词一样，说的不是同一件事。

还有一个更隐蔽的特征：**这个组织会开始特别擅长处理它见过的问题，特别不擅长处理它没见过的问题**。处理见过的问题靠流程，流程是齐的；处理没见过的问题靠判断，而判断已经不在了。于是它在稳定期表现优异，在变局中反应迟钝——而它自己会把后者解释成「转型阵痛」或者「执行不力」。

换五个场合，看同一件事

三分之一的说服力来自它在别处也咬得住。下面换五个不相干的场合，看这个结构是否还成立。这不是为了扩大适用面，是一次检验：如果它在结构相似处同样标记出同一类损失，那它刻画的就不是某个行业的特殊性。

招聘。一家公司把简历初筛交给了一套打分系统。头两年，招进来的人明显更整齐了。第五年，有人问：我们最近招的人怎么都差不多？没有人能回答，因为能回答这个问题的，是那个凭着看简历和聊天判断出「这

个人有点意思」的能力——而那个能力靠的是看过大量简历、押过几次、错过几次，这五年没有人做过这件事。

投资。一位投资人把尽调交给了外部机构。报告越来越详实，模型越来越完备。而他判断一家公司行不行的那点直觉——来自早年间自己跑现场、自己看错过、自己赔过钱——在不需要用之后，慢慢变成了对报告质量的判断力：这份报告做得够不够细。这两件事不一样，而后者不能替代前者。

医疗。一家医院把影像的初判交给了算法。放射科医生的工作变成了复核。头几年复核得很好，因为他们都是自己读片子读出来的。十年后进来的年轻医生，是在「先看算法结论再看片子」的流程里长大的——他们的眼睛被训练成检查算法有没有明显出错，而不是独立地读一张片子。当算法在某一类病例上系统性偏差时，这一代人不是不想发现，是没有那个视角。

写作与编辑。一位作者把初稿的语言打磨交给了工具。文字变干净了。而他原来那种「这一句读起来不对，我说不上来哪里不对，但要改」的感觉，在多年不被使用之后，会退化成「这一句工具没标红，应该没问题」。前者是耳朵，后者是复核清单。

家庭。父母把孩子的学业判断交给了老师和分数。这是完全合理的分工——他们确实不如老师专业。而十年后，当孩子出现了某种成绩单上看不出来的问题时，父母会说「他成绩一直挺好的呀」。这句话是真的，而且它准确地反映了他们仅有的那把尺子。

五个场合，同一个结构：**执行被交出去，判断力在若干年后无声地跟着走，而没有一处会出现异常记录。**

再补一个不成立的边界，因为边界比复现更能说明问题。在那些**结果本身会硬碰硬地反馈回来**的场合，这个结构的效力会大幅下降——比如一个人把修车交给了修理厂，车修没修好，他开两天就知道。

原因很清楚：他不需要保有判断修理工艺的能力，因为有一个**不依赖他判断的、外部的、硬的验收信号**在替他验。车会不会抛锚，不由他说了算，也不由修理厂说了算。

这个边界反过来说明了本文真正依赖的条件：**它只在「结果的好坏本身也需要有人来判断」的场合成立**。凡是好坏可以由现实直接给出、不需要人来定义的地方，交出去是安全的；凡是好坏需要有人说了才算的地方——教育、医疗的相当一部分、管理、评价、创作——交出去就带着这笔看不见的账。

而现代社会里，后一类正在快速变多。

有没有成功保住例子

到这里必须回答一个公平的质问：如果这件事这么难，那有没有人做成过？如果一个例子都举不出来，那本文说的就不是一个可对抗的机制，而是一句宿命论。

有几个例子，值得逐个看，因为它们成功的方式很不一样，而且每一个都付出了在账面上说不过去的代价。

第一个：民航的定期手飞。

飞行的自动化程度极高，而且比几乎任何行业都更早地遇到了这个问题。行业的应对是硬性规定：飞行员必须定期在模拟机上、以及在实际航段中手动操纵，即使自动驾驶完全可用。

这个做法值得注意的地方在于**它是强制的，而且是反效率的**。手飞比自动飞更耗神、更容易出小偏差、燃油经济性更差。任何一个只看单次航班效率的评估，都会得出应该取消它的结论。它之所以能维持，是因为这个行业有过足够惨痛的事故，而那些事故的调查报告把这条线写死了。

代价：它是用坠机换来的。这是最不该被浪漫化的一点——这条规矩不是靠谁想明白了才有的，是靠数字够大的事故才有的。

第二个：外科的带教。

一位主刀带一位年轻医生上台，让他做其中一段，自己在旁边。这段时间里，手术更慢、风险更高，而且如果出问题，责任在主刀。

它之所以能维持，靠的是行业内部一个不成文但很硬的共识：**不这么做，二十年后就没有会做的人了**。注意这个共识的时间尺度——它是二十年，远远超过任何一个主刀个人的考核周期。

代价：由当下这一台手术的病人，替二十年后的病人分担了一点风险。这件事在伦理上从来不轻松，各国的规范也一直在调整边界。它能成立，是因为这个代价被明确地看见并讨论过，而不是被藏起来。

第三个：军队的复盘制度。

某些军队有一套强制的行动后复盘：所有参与者到场，不论军衔，就一件事讲清楚——原计划是什么，实际发生了什么，为什么有差别，下次怎么改。关键规则是**军衔在这个房间里不生效**。

它保住的正是第三节说的第三个环节：自己发现哪里错了。而它靠的是把「说出上级判断失误」这件事的代价，用制度硬压到接近于零。

代价：极其耗时，而且它把大量指挥官的判断失误公开化——这在任何一个以权威为运转基础的组织里都是昂贵的。

三个例子有一个刺眼的共同点：**它们都不是靠理念维持的，是靠一条外部的、硬的、不容讨价还价的规定维持的**。而这条规定的来源，要么是死过人（民航），要么是行业有一个远超个人任期的时间尺度共识（外科），要么是失败的后果极其直接（军队）。

这就给出了一个不太乐观的推论：**在那些失败后果不直接、不致命、显现很慢的领域——教育、评价、管理、多数文职工作——上述三种来源全都缺席**。没有事故报告逼你，没有二十年的共识约束你，失败也不会在下一场行动里立刻惩罚你。

所以这些领域不是不能保住，是**没有任何外力会替它保住**。要保住，只能靠有人在没有任何强制、也没有任何回报的情况下，自己决定花那笔钱。

一句话的规律

把上面收拢成一句话：

你把一件事交出去的时候，交出去的不止那件事，还有你判断这件事做得好不好的能力——因为那个能力本来就是「自己做」的副产品。于是每一次委托，都同时委托掉了对这次委托的监督权。这是唯一一种你不会发现的损失：一旦发生，用来发现它的那个器官已经不在了。

短一点：

交出去的东西里，总有一样是验货的尺子。

请注意这句话的三层，缺一不可：

第一层是**验收能力是做的副产品**。这一层解释了为什么不能只交出执行、保留判断——那不是两个可以分开打包的东西。

第二层是**这个损失自我掩盖**。发现它需要用到它。这一层解释了为什么它从不以事故的形式出现，而是以「一切正常直到崩掉」的形式出现。

第三层是**修复动作由已受损的判断力主持**。这一层解释了为什么每一轮改进都精确地解决上一轮能被看见的问题，而整体一直在同一个方向上滑。

只看第一层，是「用进废退」。只看前两层，是一个聪明的悖论。三层合起来，才解释了这个过程为什么不会自我纠正，而且会随着管理精细化而加速。

与六个最像的说法划界

一个说法最容易犯的错，是把别人说过的话重讲一遍。所以要把最近的六个邻居请出来，逐一说清区别。说不清，本文就该作废。

第一个邻居：委托—代理问题。

经济学里的经典框架：委托人雇代理人办事，两者利益不一致，代理人有信息优势，于是会偷懒或谋私；解法是设计激励和监督。

区别在假设。这个框架从头到尾假设**委托人有能力判断代理人干得好不好**，只是要付出监督成本。它讨论的是成本高低、激励设计、信息不对称。

本文说的是更前面一步：**委托人验收的能力本身，会因为委托而消失**。在这个框架里，这种情况无法被表达——它会被记作「监督成本上升」，而实际发生的是监督这个动作已经失效，无论花多少钱。设计更好的激励对此无效，因为激励要生效，得有人能识别出结果的好坏。

第二个邻居：技能退化。

管理学与劳动研究里说：工作被拆解和自动化之后，从业者的技能会退化，变成只会操作某个环节的人。

这个最接近，但它讲的是**做的能力**退化。本文讲的是**验的能力**退化，而且强调的是两者的先后关系：做的能力退化通常很快被察觉（因为你一上手就发现不行了），验的能力退化不会被察觉（因为你不会去验证你的验证）。

差别的实践后果是：技能退化的常见对策是「保留一部分手工操作」「定期轮岗」。这些对保住做的能力有用，但如果轮岗时执行的是别人定好的流程、结果也不落在自己身上，那么它对验的能力没有帮助——因为第三节说的三个环节一个都没满足。

第三个邻居：自动化悖论。

航空与人因研究里的著名判断：自动化程度越高，人越少练习，于是在自动化失效、最需要人接管的那一刻，人反而最不会做。

它和本文共享同一个骨架，但它的落点是**接管时刻的操作能力**——飞机脱离自动驾驶那一瞬间，飞行员会不会飞。这是一个有明确时点、有明确失效信号的问题。

本文说的没有那个时点。判断力的失效不会给出「自动驾驶已断开」的提示音，它是连续的、无声的，而且失效之后你仍然会得到读数，仍然会做出决定，仍然觉得一切在掌握中。**自动化悖论里的人知道自己接管了；本文里的人不知道自己早就不在驾驶位上了。**

第四个邻居：古德哈特定律。

一项指标一旦成为目标，就不再是好指标——因为人会针对它优化。

区别在于**谁在动**。古德哈特讲的是被测者操纵指标，指标与真实脱钩。本文讲的是测量者失去了脱钩之后还能察觉脱钩的能力。前者是数据失真，后者是校验失灵。

一个可以同时成立的例子：一套指标完全没有被任何人操纵，数据真实无误，但它测的东西早已不是要紧的那件事——而没有人能发现，因为能发现的人都靠这套指标长大。**古德哈特描述的是被污染的水；本文描述的是尝不出水味的舌头。**

第五个邻居：外包与核心能力流失。

企业战略里的老议题：把非核心业务外包，久了会发现自己失去了对该领域的掌控，供应商反过来卡住你。

它讲的是**议价能力和依赖**，落点在权力关系上。本文讲的不是被卡脖子，而是即使供应商完全善意、完全配合、完全透明，你仍然无法判断他给你的东西好不好。**这不是权力问题，两边都没有恶意也照样发生。**

第六个邻居：知识的诅咒。

认知心理学的说法：一个人一旦懂了某件事，就很难想象不懂的人是什么感受，于是解释不清楚。

方向刚好相反，但常被混为一谈。知识的诅咒说的是懂的人的盲区，本文说的是**不再懂的人不知道自己不懂**。前者的人还有能力，只是不能换位；后者的人已经没有了那个能力，而且缺乏察觉的手段。

六条合起来：委托一代理讲监督成本，技能退化讲做不了，自动化悖论讲接管时刻，古德哈特讲指标被操纵，外包讲被卡脖子，知识的诅咒讲懂的人说不清。本文讲的是：**验的能力是做的副产品，它随着委托一起走，而且它的消失无法被它自己发现。**

如果这六条里有任何一条站不住，本文就应该退回去，承认自己只是给已有的说法换了个名字。

这不是在反对分工

必须把话说清楚，否则这篇文章会被读成「什么都要自己干」。

不是的。分工是文明最大的成果之一，而且本文所依赖的每一样东西——文字、印刷、电力、数据库——都建立在极深的分工之上。任何主张回到「凡事亲力亲为」的说法都是荒谬的：一个人一辈子也长不出所有领域的判断力，而拒绝分工的代价，是所有事都做得很差。

更要紧的是：**代理品保护弱者。**

一套统一的评分标准，对那个没有关系、不会说话、家里帮不上忙的孩子，是他唯一的指望。取消它，回到「有经验的人凭判断决定」，最先被牺牲的就是他——因为凭判断意味着凭那些人的判断，而那些人认识谁、欣赏谁，从来不是随机的。

同样，可观测的指标保护病人。一个不看化验单只凭手感的医生，遇到高手是幸运，遇到庸医是灾难，而病人事先无从分辨。

还要说清楚一点，因为它最容易被误读：**本文不是在说「专业分工让人变蠢」。**

一个只做一件事的人，在他做的那件事上，判断力会长得比什么都做的人深得多。分工不但不削弱判断力，它是判断力得以长深的前提——没有分工，没有人能在任何一件事上做够那么多次。

本文说的是另一件事：**你在某件事上交出去，你在那件事上的判断力就停了；而你并不会同时在别处长出一个能替代它的东西。**外科医生把麻醉交给麻醉师，他在麻醉上的判断力会停——这完全没问题，因为麻醉师在长。问题出在有人在长的地方：当一件事被交给一套系统、一个平台、一串指标，那件事上的判断力就没有任何人在长了。

所以真正的分界不是「交不交出去」，是**交给一个人，还是交给一套不会长判断力的东西**。前者是分工，后者才是本文说的那种委托。

所以本文的主张不是少交出去，而是：

在交出去的同时，明确地为「保住验收能力」付一笔钱。

这笔钱是纯支出，账面上永远看不出回报，而且它保护的东西在它还在的时候是隐形的。这就是它总被砍掉的原因，也是本文认为它必须被显式地写进预算、而不能指望「大家注意一下」的原因。

能做什么：六种做法与它们各自的失败模式

以下五种都有人真做过，也都有真实的失败模式。把失败模式一并写出来，是因为不写就等于卖假药。

第一种：留一条自己做的细流。

不是象征性地「体验一下」，而是真的保留一小部分完整的、自己拍板、自己担后果的事。量可以很小——一年几件——但三个环节必须齐全：决定是你做的，结果落在你身上，问题是你自己找出来的。

失败模式：这条细流会被优先安排给最闲的时候，而最闲的时候永远不来。更常见的是它变成一种仪式——领导每年下一次基层，全程被安排好，不承担任何结果。**仪式化的细流不产生判断力，它只产生做过这件事的记忆。**

第二种：用一把独立的尺子定期复核。

定期抽一小部分交出去的东西，用一种完全不依赖对方交付物的方式重新看一遍。关键是「独立」：不看他的报表、不用他的口径、不按他的分类。

比如不看学生的成绩单，而是坐在教室后面看他一节课；不看健康读数，而是问「这周你爬楼有没有比上周喘」；不看事工统计，而是找几个人聊聊他们最近在想什么。

失败模式：这件事极其耗时，产出又不可比较——你看完一节课，得到的是一个说不清的印象，而这个印象无法和任何指标对话，也无法写进任何汇报。于是它在第一次压力来临时被砍掉，理由是「效率太低」。这个理由是真的。

第三种：先自己判，再看答案。

这一条成本极低，值得单说。在拿到对方给的结论之前，先写下自己的判断——哪怕很粗糙、哪怕只有一句话。然后再看结论，比对。

它的作用不在于你判得准，而在于它保住了「**你有一个自己的判断**」这件事。一旦养成先看结论的习惯，你的判断会在你不知不觉中变成对结论的复述，而你会以为那还是你的判断。

失败模式：写下来的判断会越来越像它将要对照的那个结论——这是自然的漂移，不需要任何人施加压力。要对抗它，得偶尔在完全不知道有没有答案的情况下判一次，而这种机会需要专门制造。

第四种：把「我判不出来」作为一个合法的回答。

在任何需要判断的场合，明确允许「我没有把握」是一个正当的、不被视为无能的回答。

这一条针对的是一个具体机制：当一个人必须给出判断、而他实际上已经判不出来时，他会本能地退回到那些他还能说的东西上——格式、时限、明显的错漏。于是他给出了一个看似有内容的判断，而这个判断实际上只反映了他还剩下的那部分视力。**逼一个失去判断力的人做判断，得到的不是判断，是他视力范围的地图。**

失败模式：这个选项一旦存在，会被大量使用，包括被用来逃避。要维持它的信息量，就得追问「哪里判不出来」，而一追问，它就不再是免费的了。

第五种：把验收能力的损耗写进决策。

在决定要不要把一件事交出去的时候，除了算成本、算效率、算风险，多算一样：**交出去之后，谁还能判断这件事做得好不好，凭什么。**如果答案是「靠对方给的报告」，那就等于承认这件事从此不再受任何独立约束，这一点应当被明确地写下来，而不是默认掉。

失败模式：这一栏会被填成一句套话——「由某某部门定期评估」。要让它有意义，得追问那个部门用什么方式评估、评估的人自己做过没有。而这个追问会让原本很顺的决策流程卡住，于是它很快会被简化成一个勾选框。

第六种：让做的人和验的人偶尔换位。

针对组织。定期让负责验收的人真的去做一段那件事——不是视察，是承担一件完整的、有结果的、结果落在自己身上的活儿；同时让做的人来验一段别人的。

这一条比「轮岗」严格得多，因为它要求第三节那三个环节齐备。走过场的轮岗不算——坐在旁边看两周，不产生任何判断力。

它的好处是同时修两头：验的人重新接上现场，做的人第一次站到验的位置上，会发现自己以前交上去的东西从那个角度看是什么样。

失败模式：这件事在忙的时候第一个被取消，而且它取消得毫无痛感——因为取消的当月，各项指标都会变好看。这是所有此类做法的通病：**它们的成本立刻显现，收益要很多年后才显现，而且显现的时候没有人会把它归给这件事。**

六种做法有一个共同点：**它们的收益不可计量，成本却完全可计量。**这个不对称不会因为大家理解了它就消失。所以这六条都不是解法，只是延缓。

怎么判断这套说法是错的

一个说法如果怎么都推不翻，那它多半没说什么。以下八条，任何一条被证实，本文的判断就站不住。

第一条，关于验的能力是不是做的副产品。

找两组人：一组长期亲自做某类工作，一组长期只审阅这类工作的成果而不亲自做。给两组同样一批质量参差的作品，看谁能更准确地把好坏排序，并且看谁能指出问题所在。如果两组表现相当——那么「验是做出来的」这个判断就是错的，看得多本身就够。

第二条，关于三个环节是否都必要。

把第三节说的三个环节（自己决定、自己承担、自己发现）拆开，做四组对照：三个都有、只缺决定权、只缺后果承担、只缺自主发现。如果缺任意一个环节的组，与三个都有的组在判断力增长上没有差别——那么本文对机制的刻画就要重写。

第三条，关于损失是否真的不可自察。

对已经长期依赖某套代理的从业者，先请他们自评「我判断这件事的能力如何」，再用独立方式实测。如果自评与实测高度相关——那么「这种损失无法被自己发现」这个判断就是错的，人是知道自己退化了的。

第四条，关于改革方向的偏向。

收集若干轮评价体系改革的方案，编码它们各自解决了什么问题。如果这些方案中，指向「上一轮未被任何人抱怨过的盲区」的比例，与指向「上一轮已被抱怨的问题」的比例相当——那么「改革只解决看得见的问题」这个判断就是错的。

第五条，关于知情是否改变行为。

对一批明确知道代理代价的人，做一次充分的、有说服力的告知干预，再测其行为变化。如果告知能显著改变行为——那么第六节整个作废，这确实是个认知问题，讲道理有用。

第六条，关于四分法里的第一种是否真能维持。

找一批声称「执行交出去了但验收还在自己手上」的组织，检验其验收方式是否真的独立于对方的交付物。如果其中相当一部分确实做到了独立验收，并维持了三年以上——那么本文关于「独立验收极难维持」的判断就过于悲观。

第七条，也是最直接的一条。

如果有哪个组织建立起了一套能稳定保住验收能力的常规机制，运行五年以上没有退化成走过场——那么本文的核心判断就整个瓦解。找到反例的人不需要反驳本文的任何论证，他只需要把那个组织指出来。

第八条，关于四种形态的滑动方向。

本文说这四种只朝一个方向滑。这一条可以用历史材料检验：追踪若干个组织在十年到二十年间的验收方式变化，看有没有出现自发的反向移动——即在没有外部事故、没有强制规定的情况下，从「只看对方材料」回到「独立验收」。如果这种反向移动并不罕见，那么「不需要有人往回推也会走到第四种」这个判断就是错的。

这一条不需要做实验，只需要翻档案。

这八条都可以真去做。第一、二、三、五条是可控对照，第四、八条是文本与档案编码，第六、七条只要找反例。

那些还没来得及做过的人怎么办

前面整篇都在讲「交出去会带走已经长出来的东西」。但有一大批人面对的不是这个问题，而是另一个：**他们还没来得及长。**

一个刚入行的人，一个第一次带孩子的父母，一个进入全新领域的人——他们手上没有那张地图，也没有时间先自己踩几年坑。对他们来说，交出去不是损失，是唯一可行的起步方式：先照着标准做，先信任专业的判断，先用现成的工具。

这一点必须承认，而且它推翻了任何「凡事都该自己判断」的主张。在一个人还没有判断力的时候让他自己判断，得到的不是判断力，是随机的错误加上对自己错误判断的确信。

但这里有一个时间上的要求，而它正是今天最容易失守的地方：**这段依赖必须是有限期的。**

传统的师徒制、住院医规范化培训、飞行员的带飞小时数，都有一个共同结构：**先高度依赖，然后依赖被逐步地、强制地撤掉。**撤掉的时点是硬性规定的，不由学习者自己决定，也不由效率决定。因为如果由效率决定，那个时点永远不会到来——依赖着的人干得更稳、出错更少、更符合规范。

今天出问题的地方，往往就在于**依赖没有期限**。工具足够好用，标准足够完备，流程足够顺畅，于是「先照着做」这个阶段没有终点。一个人可以在这个阶段一直待到退休，而且待得很好。

所以对新人来说，本文的判断要反过来读：**问题不是你依赖，是没有人给你的依赖设一个到期日。**而设这个到期日必然要付出代价——放手让他自己判的那几年，事情会做得更差。这个代价在账面上很明确，收益却完全看不见。

这也解释了为什么这件事在今天特别难。过去的师徒制里，师傅有一个非常实在的动机：他会老，他需要有人接手，而接手的人得真的会。今天大量的组织没有这个时间尺度——人员流动很快，没有人指望三年后还在同一个位置上，于是没有人有动机去承担那个「让他自己判」的成本。

结果是每一代人都在最有效率的状态下工作，而没有一代人在长那张地图。

与本栏几篇的分工

本栏已有几篇都在讲「有东西在制度里悄悄没了」，容易被读成同一件事，所以要把分工说清楚。

之三《账本记不下的那样东西》做的是**记账单位**——账本记得下做了什么，记不下有人把自己搭进去。之四《为什么最该做的事，算起来总是亏的》做的是**结算时点**——好处显出来太慢，账先结完了。之十《一句对的话，也能把门关上》做的是一句**说法落到人身上之后**——它是把人推回现场，还是替他把现场关掉。之十一《没人吭声，其实有三种》做的是**读数为零的多义性**——一个位置上什么也没有时，那对应着几种不同的内部情况。

本篇做的是**委托这个动作的连带损失**：不问什么东西记不下、不问账什么时候结、不问一句话起什么作用、也不问零是什么意思，只问一件事——**当一件事被交出去之后，还有谁能判断它做得好不好，凭什么。**

前几篇的主角分别是账本、账期、说法、读数。本篇的主角是那把用来验货的尺子，以及它为什么总是和货一起被交出去。

回到那三件事

把开头三件事往后推十年。

学生和分数。十年后，那位记得「他开始会自己回头检查了」的老师退休了。接他班的年轻老师，是在完整的评价体系里长起来的——她做得非常好，各项指标都漂亮。有一天她带的班里也有一个这样的学生，而她不会注意到，因为她的注意力是被这套体系训练出来的，而这套体系里没有这一栏。

不会有人发现少了什么。少了的那样东西，上一代人从未成功地把它交出去过——他们试过，但那种判断说不清依据，写不进任何交接材料。**它不是被丢掉的，是没能被传下去的。**

信和数字。十年后，那套可观测的操作运转得更成熟了。数据更全，流程更顺，报告更漂亮。而当有人隐约觉得哪里空了，他能拿出来说的，仍然只有那些数字——他会说参与率下降了、留存不好。于是解决方案会是提高参与率。

身体和读数。十年后，那个人的读数一直很好。有一天他病了，回头想想，其实前面几个月身体一直在说些什么，只是他没有在听——不是不想听，是那套听的方式，他已经很多年没用过了。

这三处还有一个共同点，值得单拎出来：**每一处都会有人隐约觉得不对，而每一处那个人都说不出所以然。**

那位年轻老师偶尔会觉得，班上有些孩子好像不太一样，但她说不清；那个团体里有人偶尔觉得空，但他能拿出来讲的只有数字；那个人偶尔觉得身体怪怪的，但体检报告全是正常。

三处的那点隐约，都没有被当回事。因为在每一处，那点隐约都必须先被翻译成这套体系听得懂的话，才能被讨论——而它一旦被翻译，就变成了体系已经有的那些说法：这孩子成绩下滑、参与率降了、指标略高但在正常范围。**翻译的过程本身，把要说的那件事筛掉了。**

三件事的十年后指向同一件事：**没有一处会出现异常记录，因为负责记录异常的那个东西，正是被换掉的那个。**

这篇文章自己也是交出去的产物

最后必须说一句不太体面的话。

这篇文章讲的是：把事情交出去，验收的能力会跟着走。

那么这篇文章本身呢？

它是根据三篇已经写好的文章写成的。写它的人没有在教室里坐过一年，没有在那个团体里待过，没有长期依赖过健康读数然后眼看着自己的感觉退化。它所依据的每一份材料，都是别人做完之后交出来的成果。

按它自己的说法，这意味着：**写它的人没有验收这些材料的能力**。它能做的只是把三篇的判断放在一起，看它们在哪里打架。至于那三篇各自说得对不对——本文没有独立的手段去判断，它只能相信。

这不是谦虚，是这套说法的一个真实缺口。**一篇关于「不亲自做就没有判断力」的文章，本身是不亲自做写出来的**。它因此只能提出这个问题，不能示范这个问题的解法。

还有一层更实际的顾虑。这篇文章最可能的坏用法，是被拿去反对一切分工和一切标准——「你看，交出去就是有害的」。而本文第十一节已经写明，取消标准的第一批受害者是最没有话语权的人。这个误用在文字上无法被彻底挡住，任何补充的免责声明都只是一句话。

我能做的只有一件：把这一点写在这里，而不是假装没有。

如果一定要留下一句能用的，大概是这样：

下一次你把一件事交出去，顺手问一句：三年之后，我还能不看 he 给我的东西，就知道这件事做得好不好吗？如果答案是不能，那也未必就不交——但至少你知道自己这次交出去的，是两样东西。

文中提到的六个说法，出处在这里

本文不是学术论文，正文里没有加注，但把话说到了别人头上，就该把出处交代清楚，方便想深究的人自己去看。

委托—代理问题出自 Michael Jensen 与 William Meckling 一九七六年的论文，以及其后庞大的契约理论文献。

技能退化的经典讨论出自 Harry Braverman 一九七四年的《劳动与垄断资本》。

自动化悖论出自 Lisanne Bainbridge 一九八三年的短文《自动化的讽刺》，篇幅很短，值得一读。

古德哈特定律出自 Charles Goodhart 一九七五年关于货币政策的论文，后由 Marilyn Strathern 概括成流传最广的那句表述。

外包与核心能力的讨论出自 C.K. Prahalad 与 Gary Hamel 一九九〇年关于企业核心能力的论文。

知识的诅咒出自 Colin Camerer、George Loewenstein 与 Martin Weber 一九八九年的经济学实验研究。

本文与这六位的关系不是继承，也不是反驳，而是划界——他们各自说对了一件事，而本文关心的那件事，恰好落在他们六个的边界之外。如果读完之后你觉得本文说的其实就是其中某一个，那么第十节没有写成功，责任在我。

附：与本栏另外几篇的关系

本栏已有几篇都在讲「有东西在制度里悄悄没了」，容易被读成同一件事，所以要把分工说清。之三《账本记不下的那样东西》做的是**记账单位**——记得下做了什么，记不下有人把自己搭进去。之四《为什么最该做的事，算起来总是亏的》做的是**结算时点**——好处显出来太慢，账先结完了。之十《一句对的话，也能把门

关上》做的是一句说法落到人身上之后是探针还是句号。之十一《没人吭声，其实有三种》做的是**读数为零的多义性**。本篇做的是**委托这个动作的连带损失**：只问一件事——当一件事被交出去之后，还有谁能判断它做得好不好，凭什么。前几篇的主角是账本、账期、说法、读数，本篇的主角是那把用来验货的尺子，以及它为什么总是和货一起被交出去。

这一篇由哪三篇撞成

三篇对同一件事给出了方向不同的解释。同样是「明知代理会毁掉那样东西，人还是交了出去」——第一篇说是「可比较才公平」这套叙事给了它道德正当性；第二篇说不是被说服，是它偿还了一个更根本的不安：我怎么知道这是真的；第三篇说信念和需求都不是变量，因为大量的人明确知道代价，然后照做不误。三个答案指向三个互不相容的病灶：认知、需求、以及两者都不是。

教育评价·不可代理

代理如何成为暴力：高利害评价对不可代理之物的结构性排除

半个世纪所有替代代理品的改革都失败了，而人们仍认为出路是找一个更精确的代理品。

宗教社会学·制度自持

从「撑住」到「替偿」：当代基督教形态自我维持的另一种分析

把本该由不可见不可量化之物承载的确定性，改由一套可观测可管理的操作来提供——这不是阴谋，是对一种真实不安的回答。

数字代理·基底耗散

静默的荒年：数字代理依赖中「知而不动」的规范性默认机制

代理服务空前丰裕而基底能力持续耗散，且耗散并非发生在无知之中——人明确知晓代价。