arXiv:2611.30612v1 [econ.TH] 23 Nov 2026

启示原理的主要启示

Mark Whitmeyer*

2026年11月24日

摘要

我证明了揭示原理的一个逆命题，并指出它是一种乐观主义原则：若该原理必须适用于所有环境（事实上，甚至只需适用于私人价值和独立类型的环境），委托人就必须假设代理人会选择她最偏好的均衡。当委托人只知道先验分布属于某个集合时，她还必须假设实际采用的是她最偏好的先验分布。然而，在私人价值、独立类型和拟线性收益的条件下，不受限转移支付使委托人能够通过一种直接机制，逼近任何机制的均衡收益；该直接机制具有唯一均衡，且该均衡是真实的。

委托人为一组代理人设计一场博弈，即，一种机制，供他们参与。揭示原理指出，这一博弈设计问题可以大幅简化：对委托人而言，将她可能想到的任何一个供代理人参与的（同时行动）博弈替换为一种简单博弈，并不会损失最优性；在这种简单博弈中，代理人只需“报告”自己的私人信息，而新的机制本身则执行委托人可能希望实现的任何复杂行为。这一原理表明，任何博弈都可以被另一个（更加）简单的博弈所替代，同时不会损害委托人的利益。

*致 KS。

但委托人已经用一个博弈替换了另一个博弈，而不同的博弈具有不同的均衡。揭示原理成立这一事实揭示了委托人如何评价博弈——机制——以及其中进行的均衡?

本文的论点是，揭示原理是一**种乐观主义原则**：如果它普遍成立（其含义将在下文几段中作更精确的说明），委托人就必须通过假定她最偏好的均衡会被实施，来应对均衡多重性；而当她不知道环境的确切（随机）特征时，委托人就必须假定采用对她而言最有利的分布来评估其机制。

我的设定很基本。委托人与代理人对代理人的类型共享一个共同的全支持先验。委托人设计（并承诺采用）一个有限机制 M，该机制为每个代理人规定一个有限的非空消息集合 A_i 以及一个结果规则 $g\colon\ A \to \Delta(X)$（对应于一个非空的结果集合 X）。每个代理人在同时发送消息之前，都会观察机制以及自己的类型。任何此类机制都对应于一个由代理人进行的不完全信息博弈，该博弈具有相应的均衡集合，其中每个均衡都会为委托人带来一个数值收益。委托人根据函数 $V\colon \mathbf{K(R)} \to \mathbf{R}$ 评估机制；其定义域是$\mathbf{R}$的所有非空紧致子集所构成的集合，其自变量是委托人的所有均衡收益构成的集合，并且我假设该函数关于豪斯多夫距离是连续的，且其输出介于最差与最佳均衡收益之间（如果需要一个术语，可以称其为内部的 ）。

我首先证明，在每个环境中，委托人都可以在不损失最优性的情况下，将任意机制替换为真实直接机制——也就是说，揭示原理普遍成立——当且仅当$\mathcal{V}(S) = $最大值 S 对于每个 $S \in \mathcal{K}(\mathbf{R})$，也就是说，委托人必须以最乐观的方式评估任何收益多重性，并假设将实现最佳均衡。[1] 此外，即使将环境限制为具有私人

[1]"如果"当然是经典逻辑中的概念。我的贡献是逆命题。

价值和独立类型。

在我的第二个结果中，我假设委托人只知道代理人的共同先验属于某个非空紧集 P。此时，委托人首先针对 P 中的每个先验分布，根据 V（如上所述）为均衡收益集合赋予一个收益，从而评价一个机制。然后，她对 P 中所有先验分布所得到的这些收益构成的集合取闭包，并通过另一个豪斯多夫连续函数对这类集合进行评价。

在存在这一额外层次的不确定性时，要使揭示原理普遍成立，其要求出人意料地高。首先，我证明这必然要求一种我称为共同中期信念的条件；对于至少有两个代理人且每个代理人至少有两种类型的环境，这一条件要求委托人**知道代理人的共同先验**。在假设共同中期信念之后，乐观主义表现得更为强烈：两层评价都必须取最大值。

我以引入转移支付结束本文，并揭示：在委托人与代理人的收益在其中均为可加分离的环境中，委托人的评价标准几乎无关紧要。具体而言，对于由某个机制诱导的贝叶斯纳什均衡所产生的任意委托人收益，都存在一个直接机制，使其唯一均衡是真实的，并且在该机制下委托人的均衡收益可以任意接近原收益。关键洞见在于，可以修改任何严格贝叶斯激励相容直接机制中的转移支付，使配置保持不变、如实报告下的期望转移支付保持不变，同时使如实报告成为新博弈的唯一均衡。

相关文章

标准的揭示原理是，任意机制的一个均衡结果都可以作为直接机制的真实均衡结果加以复现。[2] 原机制的其他均衡不必被复现，且该原理也不会

[2]参见Dasgupta 等人（1979年）以及Myerson（1982年）。

约束直接机制中的其他均衡。完全实施则要求，机制博弈的全部均衡结果集合必须是所期望的结果。[3]

与之更为接近的是那些提出或研究设计者如何处理均衡多重性的论文。信息设计文献中有不少相关研究——其中一些要求选择设计者最偏好的均衡；另一些则要求对其最不偏好的均衡进行对抗性选择。[4] Bergemann 等(2024年赋予设计者一个定义在均衡结果集合上的任意豪斯多夫连续目标函数，并证明对于一大类信息设计问题，简单的信息结构便已足够。信息稳健机制设计主张，应根据一个机制在各种信息结构和均衡下的最差表现来对其进行评估。[5]

此外，人们还付出了大量努力，试图弱化关于代理人信念的共同知识假设。相应地，先验无关机制设计寻求在一类分布中均能表现良好的机制。在特定环境下，间接机制的表现可能优于每一种真实的先验无关机制——这种损失由揭示差距来衡量。[6] 此外，较早的研究考察了这样一种间接机制：它利用竞买人对卖方未知分布的了解。[7] 我的代理人知道先验分布，而委托人必须针对给定集合中的所有先验分布选择同一个机制。

---

[3]Palfrey and Srivastava（1989a）提供了一个配置规则的例子：尽管每个等价的直接机制都有额外的均衡结果，但某个间接机制仍能完全实施该配置规则。另见 Repullo（1985年）、Palfrey and Srivastava（1989b）、Mookherjee and Reichelstein（1990年）、Jackson（1991年）以及 Serrano and Vohra（2010年）。[4]Bergemann and Morris（2016年）和 Taneva（2019年）研究博弈中的设计者最优信息设计。Bergemann and Morris（2019年）讨论了对抗性均衡选择为何会破坏通常的直接机制化论证。Mathevet 等（2020年）允许采用不同的选择规则，并研究了对抗性选择下的一个应用；另见 Morris 等（2024年）。[5]例如，Bergemann 等（2016年）、Brooks and Du（2021年）以及 Brooks and Du（2024年）。[6]Bergemann and Morris（2005年）研究了更丰富类型空间上的机制设计。另可参阅 Chung and Ely（2007年）以及 Bergemann and Morris（2009年）。关于先验无关机制设计，参见例如 Dhangwatnotai 等（2015年）。Feng and Hartline（2018年）提出了揭示差距，而 Feng 等（2021年）则确立了收入最大化中非平凡的揭示差距。[7]Caillaud and Robert（2005年），或公共品环境下的 Choi and Kim（1999年）。

最后，我关于不受限转移的结果见§6 也与消除不良均衡的相关研究有关。在具有独立类型的拍卖中，d'Aspremont 等人(2005年)扩大竞买人的消息空间，使得每个均衡都能为卖方带来某个选定均衡下的收益。另一些论文则使用转移支付来控制直接机制中的均衡多重性。[8] 我认为，委托人可以利用转移支付来消除劣均衡，而不必假设这些均衡不会被选中。在与本文配套的一篇论文中，Whitmeyer (2026年)，我刻画了在有不完全信息和无不完全信息的博弈中，服从原则何时普遍成立。在前一种情况下，与本文一样，委托人的评价必须是最大值；在后一种情况下，评价属于一个更广泛的类别，其中同时包括最大值 和最小值。

1 草图

正如我在引言中指出的，我的两个定理的贡献在于必要性方向。对于已知先验分布，乐观主义为何必须成立，直觉上很简单。反设委托人的评价 $\mathcal{V}$ 并非总是从每个非空紧集中选出最大均衡收益，但关于豪斯多夫距离是连续的。[9] 那么，这种连续性便保证我能够找到一个有限的均衡收益集 $F=\{b,a_1,\ldots,a_k\}$，其中 $b=$最大值 $F>\mathcal{V}(F)$。

为简单起见，首先假设 $F=\{b,\ a\}$，并考虑一个包含两名代理人的环境，其中每名代理人都具有独立均匀二元类型（属于 $\{0,\ 1\}$），以及九种结果 $X=(\{0,\ 1\}\cup\{C\})^2$，其中 C 是一个额外的符号。再考虑直接机制 D^0：该机制要求每名代理人报告 0 或 1，随后在代理人选择这些报告后（$r_1,\ r_2$），选择结果 $x=(r_1,\ r_2)$。我设定代理人的收益，使得每名代理人都希望采用与另一名代理人相同的报告规则。更确切地说，每名

[8] Mathevet (2010年）和 Mathevet 和 Taneva (2013年）给出了超模机制唯一性的条件。Ollár 和 Penta (2017年、2023年）研究了在限制代理人信念的条件下如何实现完全实施，同时允许相互依赖估值，并采用较弱的共同知识假设。[9] 我尚未正式介绍这些内容，将在下一节中进行介绍。不过，跟上讨论应该并不困难。

代理人在两名代理人都如实报告或都说谎时获得收益 0，而在一名代理人如实报告、另一名代理人说谎时获得收益 -1 。

因此，代理人面临一个有两个严格（贝叶斯—纳什）均衡的协调博弈：要么双方都撒谎，要么双方都如实报告。此外，还存在一个混合策略均衡，其中每个代理人的每种类型都以相同的概率如实报告和撒谎。事实上，其余每个均衡在对每个代理人的类型进行平均后，都会使其以概率 $1/2$ 如实报告。

$(t_1,\ t_2) = (0,\ 0)$

	$r_2 = 0$	$r_2 = 1$
$r_1 = 0$	$(0,\ 0;\ b)$	$(-1,\ -1;\ \tilde{a})$
$r_1 = 1$	$(-1,\ -1;\ \tilde{a})$	$(0,\ 0;\ a)$

$(t_1, t_2) = (0,1)$

	$r_2 = 0$	$r_2 = 1$
$r_1 = 0$	$(-1,\ -1;\ \tilde{a})$	$(0,\ 0;\ b)$
$r_1 = 1$	$(0,\ 0;\ a)$	$(-1,\ -1;\ \tilde{a})$

$(t_1,\ t_2) = (1,\ 0)$

	$r_2 = 0$	$r_2 = 1$
$r_1 = 0$	$(-1,\ -1;\ \tilde{a})$	$(0,\ 0;\ a)$
$r_1 = 1$	$(0,\ 0;\ b)$	$(-1,\ -1;\ \tilde{a})$

$(t_1,\ t_2) = (1,\ 1)$

	$r_2 = 0$	$r_2 = 1$
$r_1 = 0$	$(0,\ 0;\ a)$	$(-1,\ -1;\ \tilde{a})$
$r_1 = 1$	$(-1,\ -1;\ \tilde{a})$	$(0,\ 0;\ b)$

表1：由 D^0 导出的报告博弈，其中 $\tilde{a} = (3a - b)/2$。行表示代理人1的报告，列表示代理人2的报告，约定为 $(u_1,\ u_2;\ v)$。

当两位代理人都如实报告时，我将委托人的收益设为 b ，当两位代理人都说谎时设为 a 。当一位代理人说真话而另一位说谎时，我给予委托人 $(3a - b)/2$。在涉及 C 的每个结果中，我给予委托人 a。于是，她在任何其余均衡中的期望收益为 $\frac{1}{4}b + \frac{1}{4}a + \frac{1}{2}\frac{3a-b}{2} = a$，因此她在真实均衡中获得 b，而在其他每个均衡中获得 a 。根据构造， D^0 的均衡收益集合恰好是 $\{b, \overline{a}\}$。

回到一般的有限 F，我对两个代理人分别使用具有 $k+1$ 类型和 $k+1$可用报告的相同思路，并且 D^0 再次将报告对选作结果（此时结果集合为 $X = (\{0,\ldots,k\} \cup \{C\})^2$）。代理人在观察到自己的类型后，会选择将报告相对于该类型向“前方”偏移多少，其中

这些标签按 $k+1$取模——可以把这些类型想象成围成一个圆排列。为了产生类似的协调动机，我设定代理人的收益，使得每个人都希望自己的报告相对于其类型的偏移量，等于另一位代理人的报告相对于其类型的偏移量。

因此，对于每个 j，两个代理人都使用偏移量 j 是一个严格均衡。偏移量为 0 表示如实报告；当两个代理人都使用偏移量 j时，我让委托人的收益为 a_j （因此，当他们如实报告时，即， $b=a_0$），而在涉及 C的每个结果中，委托人的收益为 a_k。我赋予代理人的二次损失收益使得，在其余每个均衡中，每个代理人的偏移量分布都以各二分之一的概率取两个连续的偏移量。正如我对双收益 F所做的那样，我选择代理人使用不同偏移量时委托人的收益，使得每个这样的混合均衡都为她带来相同数值 a_0、.....、 a_k中的一个。最终，这使得 D^0 的均衡收益集合恰好为 F。

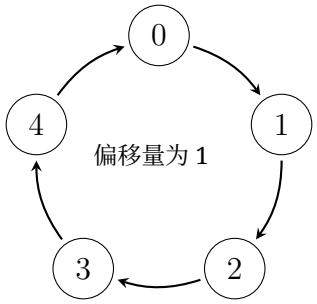


图1：五种类型下偏移量为 1：每种类型报告顺时针方向的下一个标签。更一般地，偏移量 j 使每种类型沿顺时针方向移动 j 个位置，模 $k+1$。

有了这一构造，我随后采用一种间接机制 M^+，将C 作为额外消息提供给每位代理人，并选择消息组合作为结果。我对涉及 C 的结果设定代理人的收益，使得报告 C严格支配任何虚假报告；在排除虚假报告后，如实报告又严格支配报告 C。简言之，这一间接机制使得所诱导的博弈是支配可解的，并存在唯一均衡：代理人如实报告，而委托人获得其最大收益 b。

为什么直接机制无法在消除较低收益均衡的同时也获得 b? 首先，回顾一下，D^0 不会产生涉及 C的结果。当然，直接机制可以做到这一点；但由于先验分布具有全支持，任何使委托人获得 b 的均衡，都必须在每个类型组合下将该类型组合作为结果。此外，类型组合与结果组合之间的这种精确匹配意味着，代理人的不同类型必须使用互不相交的报告集合；而由于直接机制为每位代理人提供的报告数量恰好等于类型数量，因此每种类型必须使用一个报告，均衡报告规则也必然是对类型的重新标记。结果是，机制在任何报告组合之后都无法产生涉及 C 的结果。这就确定了整个直接机制；除重新标记之外，它就是 D^0，其均衡收益集合则是已经确立的 F 。最后，我借助均衡的严格性、直接结果规则与策略组合空间的紧致性，以及V的连续性，排除了那些能够使委托人的收益任意接近 b 的直接机制。

1.1 未知先验

专门考虑单元素先验集时，第一个定理立即确定了 V 的必要形式。然而，先验不确定性还带来了另外两个问题。首先，根据某一先验下的均衡构造的直接机制，在另一先验下未必具有真实均衡。代理人根据其关于其他代理人类型的信念（以自身类型为条件），通过对偏离所带来的收益后果进行平均来评估偏离。如果委托人 P 所考虑的先验集中每个先验都诱导出相同的这类条件信念（我称之为共同中期信念），那么在每个先验下，如实报告不等式都是相同的。

共同中期信念是揭示原理成立的必要条件：如果这一条件不成立，我可以选择收益，使得同一类型在一个先验分布下偏好如实报告，而在另一个先验分布下偏好偏离。我还使委托人当且仅当结果等于类型组合时获得其最大收益。任何能够实现该收益的直接机制

随后都会被这一要求确定下来，无法规避相互冲突的如实报告不等式。因此，无论委托人如何评价由此产生的收益，都不存在任何直接机制能够在每个先验分布下诱导如实报告的同时，实现所期望的结果。

现在假设 P 具有共同中期信念。为确定对 A 的限制，固定 $a < b$，并考虑一个包含两名代理人的环境：第一名代理人只有一种可能的类型，第二名代理人有两种可能的类型。存在两种可能的先验分布 p^0，前者使类型 0 更有可能，后者使类型 1 更有可能。此外还有两个结果，即 0 和 1；当结果与代理人 2 的类型相匹配时，我赋予代理人 1 和委托人较高的收益 H，否则赋予较低的收益 L。此外，我选择 H 和 L，使得在 p^0 下，结果 0 为他们各自带来 b 的期望收益，而在 p^1 下为他们各自带来 a 的期望收益；在 p^0 下，结果 1 为他们带来 a，而在 p^1 下为他们带来 b。我使代理人 2 的两种类型都严格偏好结果 0：每种类型从结果 0 中获得效用 1，从结果 1 中获得效用 0。

现在考虑以下间接机制 M^*：代理人 1 是独裁者，可以在 0 或 1 中任选其一。因此，他在 p^0 下选择 0，在 p^1 下选择 1。结果是，无论先验分布如何，委托人都获得 b。现在反过来考虑任意一种直接机制，在两种先验分布下代理人都如实地报告。该机制无法根据代理人 2 的信息（即他对自身类型的了解）进行条件化，因为他的两种类型都严格偏好结果 0。但代理人 1 的报告毫无用处——除先验分布之外，他没有其他私人信息，而先验分布不属于机制的内容。[10] 因此，尽管我们为委托人在两种p下都选择最佳均衡，但如果委托人的先验特定值集合包含 b，那么它必定是 $\{a, b\}$；因此，依据 A 的连续性（和内含性），必有 $\mathsf{A}(\{a, b\}) = b$。最后，这一逻辑很容易推广到有限个 F，而 A 的连续性则可进一步将其推广到紧集。

[10]如果将先验分布也纳入代理人的类型，揭示原理再次等价于代理人所进行的均衡上的乐观主义。参见 §3.1。

1.2 转移支付使评价标准失去意义

或许令人意外的是，当委托人可以自由选择货币转移支付，且委托人与代理人的效用关于转移支付是可加分离的时，评价标准变得无关紧要；也就是说，存在一种真实直接机制，使委托人在博弈的唯一均衡中获得的期望收益达到任意接近。

关键思想是：在任何严格贝叶斯激励相容（BIC）直接机制中，我都可以修改代理人的支付，使得当其他代理人如实报告时，每个代理人从任何报告中获得的期望支付保持不变；但当其他代理人不如实报告时，支付可以有所不同，从而消除所有其他均衡。转移支付变换如下进行。我从严格贝叶斯激励相容直接机制开始，对每个代理人（假设有多个代理人的类型数大于一个，否则问题已经解决），选择一个报告作为参考报告，并将其他每个报告分别与另一名代理人的一个报告配对，同时确保每个报告至多使用一次。如果仍有报告未配对，则这些报告必然全部属于同一名代理人，我便将该代理人单独列出。如果没有未配对的报告，我就任选一名类型数大于一个的代理人单独列出。对于每一组这样的配对，我从每个报告的指标函数中减去如实报告产生该报告的概率，再将所得的两个项相乘，把乘积加到一名代理人的支付中，并从另一名代理人的支付中扣除。对于某名代理人的每个报告，当其对手如实报告时，这种支付调整的期望值为零，因此如实报告与其他报告之间的所有比较均保持不变，如实报告仍然是严格均衡。在进行下面的缩放之前，对于每个报告组合，这些调整也满足预算中性。

为了了解这些调整在如实报告之外实现了什么，暂时忽略原机制，让代理人只关心这些调整。每位代理人都可以按照相应类型的事前概率选择每份报告，从而获得零收益。考虑这一仅调整博弈的任意均衡。由于每位代理人都可以获得零收益，因此每位代理人的均衡收益至少为零。由于

在每个报告组合下，代理人的收益总和为零，因此每位代理人的均衡收益必定等于零。面对对手的均衡策略，没有任何报告能使代理人获得高于零的收益；然而，按照事前（全支持）概率选择报告会产生零期望收益，因此面对对手的均衡策略，每份报告都必须带来零期望收益。

现在考虑我在构造中配对的两份报告。将一名代理人从其配对报告中获得的收益与其从参考报告中获得的收益进行比较，可以看出另一份配对报告必须以相应类型的事前概率被使用，而这恰好是如实报告会生成该报告的概率。对另一名代理人做同样的比较，则表明第一份配对报告也必须以其所对应类型的事前概率被使用。对所有配对重复这些比较即可表明，在这一仅调整博弈的每一个均衡中，除我单独指出的那一名代理人外，每一名代理人都以相应类型的事前概率使用每一份报告。

回到修改后的机制所诱导的博弈，接下来我缩放支付调整，使其对被单独选出的代理人仅线性增长，而对其余代理人则按二次方增长。反设对于任意大的调整规模，都存在一个非真实均衡，于是我可以选取一列趋于无穷大的调整规模，并在每个规模下选取一个非真实均衡。原机制中的收益有界。此外，对于每个代理人和每一种可能的报告，由于我们处于均衡，任何代理人的任何类型都不存在有利偏离。我对这些不等式按代理人的类型取平均，再除以其支付调整的规模。由于原机制中的收益有界，随着规模增大，它们对激励的影响会趋于消失。因此，代理人报告分布的任何极限都必定是仅由这些调整构成的博弈的均衡。于是，沿着假定存在的均衡序列，每个这样的代理人使用各个报告的概率都收敛于相应类型的事前概率。

为什么剩下的代理人也必须这样做？一旦其对手的报告分布接近如实报告所产生的分布，严格的贝叶斯激励相容性就会使每个虚假报告都严格劣于真实报告。因此，他完全转而如实报告所获得的收益，至少等于一个正数乘以其误报概率。另一方面，我可以对代理人的均衡不等式加权并求和，使支付调整相互抵消。由于其他代理人的调整按二次方增长，而他的调整仅线性增长，所得不等式会将他的收益乘以共同的规模，同时使其他代理人的收益不受缩放。那些其他代理人的收益有界，因此他的收益以及他的误报概率都必定收敛于零。

因此，所有代理人的报告分布都接近如实报告所产生的分布。严格贝叶斯激励相容性现在使每个代理人完全转向如实报告所获得的收益都为正；只要该代理人以严格正的概率作出不实报告，这一收益就严格为正。然而，如果我将每个表述“代理人在修改后的机制中无法通过转向如实报告而获利”的不等式，除以乘以其支付调整额的系数，然后把这些不等式相加，调整项就会相互抵消。所得的加权和，即各代理人在原机制中转向如实报告所获得的收益之和，必定是非正的。由于这些收益中的每一个都为正，因此它们必定全部等于零；于是没有代理人作出不实报告，这与假定存在另一个均衡相矛盾。

2 设计问题

有一位委托人（女性）和一个由代理人组成的有限集$\{1,\dots,n\}$，其中有 $n \geq 1$。代理人i 私下观察其类型 t_i ，该类型属于一个有限非空集 T_i。类型组合 $t = (t_1,\dots,t_n)$ 根据具有全支持的共同先验 $p \in \Delta(T)^\circ$ 抽取，[11] 其中 $T := \prod_i T_i$，而 $p(\cdot \mid t_i)$ 表示给定 t_i时其他代理人类型的条件分布。

[11] $\Delta(Y)$ 表示有限非空集合 Y 上的概率分布集合，其中 $\Delta(Y)^\circ$ 表示具有全支持的概率分布。

如果对于每个 $t \in T$都有 $p(t) = \prod_i p_i(t_i)$，则代理人的类型是独立的，其中 p_i 是 p 在 T_i上的边际分布。

存在一个有限的非空结果集合 X。代理人 i 从结果 x 和类型组合 t中获得效用 $u_i(x,t) \in \mathbb{R}$，而委托人获得收益 $v(x,t) \in \mathbb{R}$。所有参与方都根据期望效用评价结果上的彩票：对于 $\ell \in \Delta(X)$，记作$u_i(\ell,t) := \sum_x \ell(x)u_i(x,t)$ 和 $v(\ell,t) := \sum_x \ell(x)v(x,t)$。在我的主要设定中，类型集合 $(T_i)_{i=1}^n$、结果集合 X、先验分布 p以及支付函数 $(u_i)_{i=1}^n$ 和 v 都是共同知识。如果每个代理人的效用仅通过其自身类型取决于真实类型组合，则称环境具有私人价值；在这种情况下，我将代理人的效用写作 $u_i(x,t_i)$。

2.1 主要设定

在类型抽取之前，委托人承诺采用一个机制。一个有限机制 M 为每个代理人指定一个有限的非空消息集合 A_i 以及一个结果规则 $g\colon A \to \Delta(X)$，其中 $A := \prod_i A_i$。每个代理人在同时发送消息之前，都会观察机制和自己的类型。随后，根据 $g(a)$抽取结果，其中 a 是消息组合。

策略 $\sigma_i\colon T_i \to \Delta(A_i)$ 指定了代理人 i 在每种类型下对消息的分布。用 $\sigma = (\sigma_1, \dots, \sigma_n)$ 表示一个策略组合。我用 t_{-i}、a_{-i} 和 σ_{-i} 分别表示其他代理人的类型、消息和策略。对于类型为 t_i 且发送消息 a_i的代理人，当其他代理人遵循 σ_{-i} 时，记其期望效用为

$$U_i(a_i \mid t_i; M, \sigma_{-i}) := \sum_{t_{-i}} p(t_{-i} \mid t_i) \sum_{a_{-i}} \left(\prod_{j \neq i} \sigma_j(a_j \mid t_j) \right) u_i\left(g(a_i, a_{-i}), t\right).$$

定义 2.1。一个贝叶斯纳什均衡（*BNE*）的 M 是一个策略组合 σ 满足

$$\sum_{a_i \in A_i} \sigma_i(a_i \mid t_i) U_i(a_i \mid t_i; M, \sigma_{-i}) \geq U_i(a_i' \mid t_i; M, \sigma_{-i}) \tag{1}$$

对于每个代理人 i、每种类型 $t_i \in T_i$ 和每条消息 $a_i' \in A_i$，我用 $\mathrm{Eq}(M)$ 表示所有贝叶斯纳什均衡的集合。

如果每个代理人的每种类型都以概率 1 选择一条消息，则该均衡为纯均衡。如果每种类型在固定其他代理人策略的情况下，都严格偏好为其指定的消息，而非任何其他消息，则纯均衡为严格均衡。

在策略组合 σ 下，抽取类型之前委托人的期望收益为

$$\Pi(M,\sigma) \coloneqq \sum_{t\in T} p(t) \sum_{a\in A} \left(\prod_i \sigma_i(a_i \mid t_i) \right) v(g(a),t). \tag{2}$$

委托人面对的是整个均衡收益集合 $S(M) \coloneqq \{\Pi(M,\sigma) : \sigma \in \mathrm{Eq}(M)\}$，该集合非空且紧致。

若对每个代理人都有 $A_i = T_i$，则称机制为直接的。我将直接机制D 与其结果规则 $D\colon T \to \Delta(X)$ 等同起来。若每个代理人始终报告自身类型所构成的策略组合是一个均衡，则称该机制真实的。

令 $\mathcal{K}(\mathbf{R})$ 表示 $\mathbf{R}$ 的非空紧致子集族。委托人根据 M，通过 $\mathcal{V}(S(M))$ 对机制进行评价，其中 $\mathcal{V}\colon \mathcal{K}(\mathbf{R}) \to \mathbf{R}$。[12] 我始终假设，该评价介于可能收益的最差值与最佳值之间：

$$\min S \le \mathcal{V}(S) \le \max S \qquad \text{for every } S \in \mathcal{K}(\mathbf{R}). \tag{3}$$

将这样的 V 称为内在的，并注意到这意味着 $\mathrm{V}(\{s\}) = s$。我还假设 V 关于豪斯多夫距离连续：[13] $d_H(S_m, S) \to 0$ 意味着 $\mathrm{V}(S_m) \to \mathrm{V}(S)$。

对于一个环境 E，令 $\mathrm{M}(E)$ 表示有限机制的集合，$\mathrm{D}(E)$ 表示其直接机制，$\mathrm{D}^T(E)$ 表示其真实直接机制。委托人的无限制价值，以及她在被限制为真实直接机制时的价值分别为

$$V(E) \coloneqq \sup_{M\in\mathcal{M}(E)} \mathcal{V}(S(M)), \qquad \& \qquad V^T(E) \coloneqq \sup_{D\in\mathcal{D}^T(E)} \mathcal{V}(S(D)). \tag{4}$$

常数结果规则是真实的，因此这些机制集合非空。

---

[12] Bergemann 等人 (2024年) 也在信息设计问题中，将设计目标表述为均衡结果集合的连续函数。这里 $\mathcal{V}$ 的自变量是委托人的标量期望收益集合。

[13] 对于 $x \in \mathbf{R}$ 和 $S, L \in \mathrm{K}(\mathbf{R})$，定义 $\mathrm{dist}(x, S) \coloneqq$ 最小值$_{s\in S} |x - s|$ 以及 $d_H(S, L) \coloneqq$最大值 {最大值$_{s\in S}\,\mathrm{dist}(s, L)$，最大值$_{\ell\in L}\,\mathrm{dist}(\ell, S)$}。当任一集合中的每个收益都接近另一个集合中的某个收益时，这两个集合在该距离下就是接近的。

2.2 未知先验

我还将通过以下假设扩展之前的设定：委托人不知道代理人共同拥有的先验分布。对于不含先验分布的类型集、结果和支付函数，记作 $E^0 := ((T_i)_{i=1}^n,\ X,\ (u_i)_{i=1}^n,\ v)$。令 P 为 $T := \prod_i T_i$ 上全支持分布构成的非空紧集。委托人知道 E^0 和 P，并承诺采用一种机制。随后，自然从 P 中选取一个先验分布 $p \in$ ，该先验分布是代理人之间的共同知识。然后按照 p 抽取类型，各代理人在机制运行前观察自己的类型。

委托人首先针对每个先验分布评估均衡多重性：$c_p(M) := \mathrm{V}(S_p(M))$。然后，对于有限机制 M，定义紧集$C_{\mathscr{P}}(M) := \mathrm{cl}\,\{c_p(M) :\ p \in \mathrm{P}\}$。接下来，存在第二个泛函 A：$\mathrm{K}(\mathbf{R}) \to \mathbf{R}$，委托人通过它评估关于先验分布的不确定性，因此她对 M 的总体评价为 $\mathrm{A}(C_{\mathscr{P}}(M))$。简而言之，V 描述委托人如何处理给定先验分布条件下的均衡多重性，而 A 描述她如何处理由可能的先验分布所生成的价值集合。

我假设 A 关于豪斯多夫距离是连续的，并且具有内部性：

$$\min K \le \mathcal{A}(K) \le \max K \qquad \text{for every } K \in \mathcal{K}(\mathbf{R}). \tag{5}$$

与 V 一样，$\mathrm{A}(\{k\}) = k$。

接下来我 陈述一个关于代理人在得知自身类型后的信念的条件。

定义 2.2。集合 P 具有共同中期信念，当且仅当对于每个 $p,\ q \in$ P、每个至少具有两种类型的代理人 i 、每个 $t_i \in T_i$ 以及每个$t_{-i} \in T_{-i}$，都有 $p(t_{-i} \mid t_i) = q(t_{-i} \mid t_i)$。

如果在每个 $p \in$ P 下，如实报告都是均衡，则称直接机制对于（E^0，P）是稳健真实的。用 $\mathrm{D}^R(E^0,\ \mathrm{P})$ 表示这一类机制。无约束的和稳健真实的

真实价值为

$$W(E^0,\mathscr{P}) \coloneqq \sup_{M\in\mathcal{M}(E^0)} \mathcal{A}(C_{\mathscr{P}}(M)), \qquad W^R(E^0,\mathscr{P}) \coloneqq \sup_{D\in\mathcal{D}^R(E^0,\mathscr{P})} \mathcal{A}(C_{\mathscr{P}}(D)).$$

3 特征刻画

首先，将机制限制为真实直接机制，当且仅当委托人在每个环境中都根据其最优均衡评价每个机制时，这一限制才会保留委托人的最优值：

定理 3.1。 以下条件等价：

(i) $V^T(E) = V(E)$ 对于每个有限环境 E。(ii) 在每个有限环境 E中，对于每个 $M \in \mathcal{M}(E)$ ，存在一个 $D \in \mathcal{D}^T(E)$，使得 $\mathcal{V}(S(D)) \geq \mathcal{V}(S(M))$。

(iii) $\mathcal{V}(S)$ = 最大值 S 对于每个 $S \in \mathcal{K}(\mathbf{R})$。

如果这些条件不成立，则存在一个环境，其中有两个代理人，其类型相互独立且服从均匀分布；并且存在一个有限机制 M^+ ，该机制具有唯一均衡，使得

$$\sup_{D\in\mathcal{D}(E)} \mathcal{V}(S(D)) < \mathcal{V}(S(M^+)) = V(E). \tag{6}$$

当条件(i) 和(ii) 仅要求对具有私人价值和独立类型的环境成立时，这一等价性同样成立。若这些条件不成立，则使用两个代理人、私人价值和独立类型，也可以得到(6) 。

放弃委托人知道先验分布这一假设后，我得到第二个定理：

定理 3.2。 假设$\mathcal{A}$ 是连续的，并满足(5)。以下条件等价：

(i) $W^R(E^0, \mathscr{P}) = W(E^0, \mathscr{P})$ 对于每个有限环境 E^0 以及每个由具有共同中间信念的全支持先验构成的非空紧集$\mathscr{P}$ 。

(ii) 对于每个这样的$(E^0,\mathscr{P})$ 以及每个 $M \in \mathcal{M}(E^0)$，存在一个 $D \in \mathcal{D}^R(E^0,\mathscr{P})$，使得 $\mathcal{A}(C_{\mathscr{P}}(D)) \geq \mathcal{A}(C_{\mathscr{P}}(M))$。

(iii) 对每个 S，V(S) = 的最大值为 S，且A(K) = 的最大值为 K。$K \in$ K(R)。

假设 $\mathcal{V}(S)$ = 最大值 S，且 $\mathcal{A}(K)$ = 最大值 K。对于任意非空紧集 $\mathscr{P}$，其中包含全支持先验但不具有共同中间信念，存在一个有限环境 E^0，它具有相同的类型集和私人价值，使得

$$W^R(E^0,\mathscr{P}) < W(E^0,\mathscr{P}) = 1. \tag{7}$$

在已知先验分布的情况下，直接机制化可以再现任意选定的均衡，而问题在于委托人如何评估直接机制产生的额外均衡。在先验分布未知的情况下，同一个直接规则还必须满足由每一种可能的中间信念体系所产生的如实报告不等式。但共同中间信念会使这些不等式在不同先验分布下完全相同。另一方面，一旦这一条件不成立，**任何**关于均衡多重性或先验不确定性的态度，都无法恢复普遍适用的揭示原理。

推论 3.3。 假设至少有两个代理人各自至少有两种类型。一组全支持先验只有在其为单元素集合时，才具有共同中间信念。

证明。 固定 p，$q \in$ P。对于每个 i 至少有两种类型的代理人，共同中期信念意味着 $p(t)\,q(t) = p_i(t_i)\,q_i(t_i)$ 对于每个 $t \in T$均成立。如果代理人 i 和 j 都至少有两种类型，那么似然比 $p(t)/q(t)$ 因此既仅仅是 t_i 的函数，也是仅仅是 t_j的函数。由于每个类型组合的概率都严格为正，该比值必须在 T 上保持不变。两个先验分布的总质量均为一，因此该常数为一，且 $p = q$。■

当至少两位代理人各自至少有两种类型时，任何非单点的全支持先验紧致集都对应着一个满足（7）的私人价值环境，即使两种评估都选择最大值。

3.1 更丰富的报告

模糊先验结果当然在一定程度上取决于“类型”的定义以及报告类型的含义。具体而言，尽管代理人观察到实现的先验分布 p，直接机制却只要求代理人 i 报告 t_i。如果现在允许直接机制征询每位代理人的全部中间信息（p，t_i），会怎样呢?

固定一个由全支持先验构成的非空紧集 P。由于每个代理人在发送消息之前都会观察p，因此，由机制诱导的博弈的均衡是一个可测族 $\sigma=(\sigma^p)_{p\in\mathscr{P}}$，使得对于每个 $p\in\mathrm{P}$，都有 $\sigma^p\in\mathrm{Eq}^p(M)$。令 $\mathrm{Eq}^{\mathscr{P}}(M)$ 表示这些均衡的集合，并且每个均衡 $\sigma\in\mathrm{Eq}^{\mathscr{P}}(M)$ 都会为委托人产生先验分布相关的收益 $\pi(M,\ \sigma)(p):=\Pi_p(M,\ \sigma^p)$ $(p\in\mathrm{P})$。

对于每个 $\mathscr{P}$，取一个递增函数 $\rho_{\mathscr{P}}$，该函数为每个有界可测函数 $f\colon\ \mathscr{P}\to\mathbf{R}$。[14] 我还假设它对单元素先验集进行了归一化：$\rho_{\{p\}}(f)=f(p)$。由该博弈的均衡为委托人产生的价值集合为 $S^{\rho\,\mathscr{P}}(M):=\mathrm{cl}\,\{\rho_{\mathscr{P}}(\pi(M,\ \sigma))\colon\ \sigma\in\mathrm{Eq}^{\mathscr{P}}(M)\}$，委托人根据 V $(S^{\rho\,\mathscr{P}}(M))$ M 对其进行评估。

在本小节中，一个直接机制 赋予代理人 i 消息空间 P $\times\,T_i$，并具有可测的结果规则。当在每个实现的先验 p下，报告 $(p,\ t_i)$ 都构成均衡时，该机制是真实的 。

命题 3.4。 设$(\rho_{\mathscr{P}})_{\mathscr{P}}$ 对于单元素先验集是递增且规范化的。

以下各项等价：

(i) 对于每个有限环境 E^0、每个全支持先验的非空紧集$\mathscr{P}$ 以及每个有限机制 M，都存在一个真实直接机制 D，使得 $\mathcal{V}\,(S^{\rho\,\mathscr{P}}(D))\geq\mathcal{V}\,(S^{\rho\,\mathscr{P}}(M))$。(ii) $\mathcal{V}(S)$ = 最大值 S 对于每个 $S\in\mathcal{K}(\mathbf{R})$。

证明。 必要性 最大值 显然可由 定理 3.1 令 $\mathscr{P}=\{p\}$即可得到。充分性是经典揭示原理的一个直接推论。 ■

[14]所谓递增，是指对于每个 $p\in\mathrm{P}\implies\rho_{\mathscr{P}}(f)\leq\rho_{\mathscr{P}}(h),\ \ f(p)\leq h(p)$

4 已知先验的证明

对于任何包含至少两个委托人收益的有限集 F，我构造一个环境，其中真实直接机制的均衡收益集合恰好为 F。为每位代理人增加一条额外消息即可消除劣均衡。此外，任何能够实现最佳收益的其他直接机制都必须重现整个集合 F。严格均衡随后保证，接近该最佳收益的直接机制也具有收益接近 F 中每个元素的均衡。[15]

直接机制 D 的重新标记会重新排列每个代理人报告的标签。形式化地说，对于双射 $\pi_i\colon\ T_i \to T_i$，重新标记后的机制将彩票$D(\pi_1^{-1}(r_1).....,\ \pi_n^{-1}(r_n))$分配给报告组合 $(r_1,\ .....,\ r_n)$。

引理 4.1。令 $F = \{a_0, a_1, ..., a_k\}$，其中 $b\colon = a_0 > a_1 > \cdots > a_k$ 且 $k \geq 1$。存在一个包含两个代理人和独立均匀类型的环境、一个真实直接机制 D^0，以及一个为每个代理人增加一条额外消息的机制 M^+，满足：

（i）对于每个类型组合，委托人的收益至多为 b，且恰有一个结果使她获得 b。这些最大化结果在不同类型组合之间各不相同。（ii）$S(D^0) = F$，并且 F 的每个成员都在D^0的严格纯策略均衡中实现。（iii）M^+具有唯一的贝叶斯纳什均衡，其中 $S(M^+) = \{b\}$。（iv）任何满足 $b \in S(D)$ 的直接机制 D 都是 D^0 的重新标记。

本引理的证明见附录 A.1，其中我详细阐述了§1中的构造。由此可知，任何满足 D 及 $b \in S(D)$ 的直接机制都是 D^0的重新标记，因此 $S(D) = F$。下一个引理（证明见附录 A.2）还对最佳均衡收益仅接近 b的直接机制作出约束。这一点很有必要，因为（4）中的价值是上确界。

[15] Lehrer 等人（2011年）研究哪些集合可以作为有限博弈的均衡收益集合出现。

引理 4.2。在引理 *4.1*的环境中，对于每个 $\eta > 0$ ，都存在一个 $\delta > 0$，使得每个直接机制 D 都满足

$$\max S(D) > b - \delta \quad \Longrightarrow \quad d_H(S(D), F) < \eta. \tag{8}$$

引理 4.3。 对于每个有限机制 M 以及每个 $\sigma \in \mathsf{Eq}(M)$，都存在一个真实直接机制，使其真实均衡在每个类型组合下诱导出与 (M, σ) 相同的结果分布。

证明。 这是标准的揭示原理构造，源自Myerson(1982年)。 ■

4.1 定理 3.1的证明

定理 3.1 的证明。首先，假设对于每个紧集，$\mathrm{V}(S)$ = 最大值 S 。固定M并选择一个达到最大值 $S(M)$ 的均衡。根据引理 4.3，存在一个真实直接机制 D ，其真实均衡为委托人带来相同的收益。因此，$\mathrm{V}(S(D))$ =最大值 $S(D)$ ≥最大值 $S(M)$ =$\mathrm{V}(S(M))$，这证明了 (iii) 蕴含 (ii)。(ii) 通过取上确界并利用 $\mathrm{D}^T(E) \subseteq \mathrm{M}(E)$ 蕴含 (i)。

为证明必要性，假设对于某个 $S \in \mathrm{K}(\mathbf{R})$，$\mathrm{V}(S)$ <最大值 S，并记 b =最大值 S。选择包含 b 的有限子集 $F_m \subseteq S$，使得 $d_H(F_m, S) \to 0$。[16] V的连续性意味着 $\mathrm{V}(F_m) \to \mathrm{V}(S) < b$。固定一个满足 $\mathrm{V}(F) < b$ 的有限集 F 。由于 $\mathrm{V}(\{b\}) = b$，该集合至少包含两个成员。

使用 引理 4.1 中的环境和机制 F。令 $\gamma := (b - \mathcal{V}(F))/2 > 0$。由于 $\mathcal{V}$ 在 F 处具有连续性，存在一个 $\eta > 0$ 使得

$$d_H(L, F) < \eta \quad \Longrightarrow \quad \mathcal{V}(L) < b - \gamma. \tag{9}$$

令 $\delta > 0$ 由引理 4.2 给出 η。每个直接机制 D 都属于以下两种情况之一。若最大值 $S(D) \leq b - \delta$，则由（3）可得$\mathcal{V}(S(D)) \leq b - \delta$。若

[16]由于 S 是紧致的，这样的集合确实存在：对于每个 m，取有限多个 S 中的点，使其与 S中的每个点的距离小于$1/m$ ，然后加入 b。

最大值 $S(D) > b - \delta$，则由（8）和（9）可得$\mathcal{V}(S(D)) < b - \gamma$。因此，

$$\sup_{D \in \mathcal{D}(E)} \mathcal{V}(S(D)) \leq b - \min\{\delta, \gamma\} < b. \tag{10}$$

委托人的收益 $v(x, t)$ 不会超过 b，而 M^+ 具有单元素收益集$\{b\}$。因此，$V(E) = \mathcal{V}(S(M^+)) = b$。(10) 证明了 (6)，并表明 (i) 不成立。关于私人价值的论断证明见 §B。 ■

5 未知先验证明

引理 5.1。 设有一个有限环境 E^0。对于每个有限机制 M，函数 $p \mapsto$ 最大值 $S_p(M)$ 在全支持先验的集合上是上半连续的。

我在 附录 A.3中证明该引理。

引理 5.2。 假设 $\mathcal{P}$ 具有共同中期信念。固定 $p^* \in \mathcal{P}$、一个具有结果规则 g 的机制 M，以及一个均衡 $\sigma^* \in \mathsf{Eq}^{p^*}(M)$。直接机制$D(x, r) := \sum_{a \in A} \left(\prod_i \sigma_i^*(a_i | r_i)\right) g(a(x))$ $(x \in X, r \in T)$ 在每个 $p \in \mathcal{P}$ 下都是真实的。

我在附录 A.4中证明了这一点。接下来的两个引理指出了委托人对先验不确定性的评价，以及共同中间信念的必要性，并再次补全了§1中存在的缺漏。参见附录 A.5 和 A.6。

引理 5.3。 令 $F = \{b, a_1, \ldots, a_k\}$，其中 $k \geq 1$，$b =$最大值 F，且对每个j都有 $a_j < b$。存在一个有限环境 E^0 和一个由具有共同中间信念的全支持先验构成的有限集 $\mathcal{P}$，使得对每个内部$\mathcal{V}$以及每个连续的内部$\mathcal{A}$，以下陈述均成立。

(i) 最优非约束值等于 b。(ii) 每个稳健真实直接机制都会针对某个 $q \in [0, 1]$ 产生先验特定值集合

$$K_q := \{qb + (1 - q)a_j,\ qa_j + (1 - q)b \colon j = 1, \ldots, k\}. \tag{11}$$

Co反过来，每个 $q \in [0,1]$ 都由稳健真实直接机制实现 m.

引理 5.4。固定一个不具有共同中期信念的全支持先验的非空紧集$\mathscr{P}$ 。存在一个具有相同类型集和私人价值的有限环境 E^0 以及一个数值 $\bar{v} < 1$ ，使得当 $\mathcal{V}(S)$ $=$ 最大值 S 且$\mathcal{A}(K)$ $=$最大值 K， $W(E^0, \mathscr{P}) = 1$，而 $W^R(E^0, \mathscr{P}) \leq \bar{v}$。

定理 3.2 的证明。我先证明充分性。假设对于每个紧集，V(S) $=$ 最大值 S 且 A(K) $=$ 最大值 K。固定满足共同中期信念的 (E^0, P)，并固定 $M \in$ M(E^0)。委托人赋予 M 值 A($C_{\mathscr{P}}(M)$) $=$最大值$_{p \in \mathscr{P}}$ 最大值 $S_p(M)$。根据引理 5.1以及P的紧致性，某个 $p^* \in$ P取得该最大值。选择一个均衡 $\sigma^* \in$ Eq$_{p^*}$(M)，取得最大值 $S_{p^*}(M)$。根据引理 5.2，从 σ^* 获得的直接机制 D 在P中的每个先验分布下都是真实的。其在 p^* 下的真实均衡为委托人带来最大值 $S_{p^*}(M)$。由于V和A都选择最大值，A($C_{\mathscr{P}}$(D)) $\geq$ A($C_{\mathscr{P}}(M)$)。

接下来我证明必要性。先验分布的单元素集满足共同中期信念，而（5）蕴含 $\mathcal{A}(\{c\}) = c$。因此，定理 3.1 中的等价关系意味着，普遍稳健揭示要求 $\mathcal{V}(S) =$ 最大值 S 对每个紧集都成立。

接下来需要确定 A。固定一个有限集 $F = \{b, a_1, \ldots, a_k\}$，其中 $b =$ max F，$a_j < b$，且 $k \geq 1$，并应用引理 5.3。无限制价值为 b。假设 A(F) $< b$。当 $q = 0$ 或 $q = 1$时，(11) 中的集合等于 F，因此委托人赋予它的价值低于 b。对于 $q \in$ (0, 1)，所有先验特定值都严格低于 b，所以由 (5) 可知，她再次赋予该集合低于 b的价值。由于 A 连续，她的评估随 q 连续变化，并在 $[0,1]$上达到最大值。这个最大值严格低于 b，因为她在每个 q处的评估都低于 b。因此，无限制价值与稳健真实价值相等，要求对每个至少包含两个元素的有限集 F ，A(F) $=$max F。对于单元素集合，由 (5) 可得相同结论。对于每个 $K \in$ K(R)，选取有限集 $F_m \subseteq K$，使得 $d_H(F_m, K) \to 0$。由 A 的连续性以及最大值的连续性可得 A(K) $=$ lim$_m$ A(F_m) $=$lim$_m$ max $F_m =$max K。这就证明了等价性。

最后，引理 5.4 对每个违反共同中期信念条件的7) 紧致全支持先验集合给出结论。 ■

6 不受限转移

现在我允许委托人自由选择货币转移支付。正如我们将看到的，揭示原理不再要求乐观主义。保留有限类型集 T_i 和私人价值，并假设在已知的全支持先验分布 p下，代理人的类型相互独立。在结果 x、类型组合 t以及代理人支付的支付 $\tau = (\tau_1, \ldots\ldots, \tau_n)$下，代理人 i 获得效用 $u_i(x, t_i) - \tau_i$，委托人则获得$v(x, t) + \sum_i \tau_i$。我假设，对于每个代理人 i及任意两个不同的类型 $t_i, s_i \in T_i$，函数 $x \mapsto u_i(x, t_i) - u_i(x, s_i)$ 在 X 上不是常数。

在本节中，有限机制 $M = ((A_i)_i, q, \tau)$ 规定了有限的非空消息集、一个分配规则 $q\colon A \to \Delta(X)$，以及支付 $\tau_i\colon A \to \mathbf{R}$。直接机制对每个代理人都设有 $A_i = T_i$。用 σ^T 表示如实报告；如果在其他代理人如实报告时，每种类型都严格偏好如实报告而非任何其他报告，就称该直接机制严格贝叶斯激励相容（严格 *BIC*）。

对于策略组合 σ，令 $q^\sigma(x \mid t)$ 表示在真实类型组合 t 下结果 x 发生的概率，并令 $\bar{\tau}_i^\sigma(t_i)$ 表示类型 t_i 的中期预期支付。用 $U_i(t_i; M, \sigma)$ 表示其期望净效用。委托人的预期利润为

$$\Pi(M, \sigma) = \sum_{t \in T} p(t) \sum_{x \in X} q^\sigma(x \mid t) v(x, t) + \sum_i \sum_{t_i \in T_i} p_i(t_i) \bar{\tau}_i^\sigma(t_i).$$

与之前一样，公式（M）包含所有贝叶斯纳什均衡，且 $S(M) = \{\Pi(M, \sigma)\colon \sigma \in \text{公式（}M\text{）}\}$.

代理人具有依类型而定的外部选项：$o_i(t_i) \in \mathbf{R}$ 是类型 t_i 的外部选项（以效用计）。如果一个均衡满足以下条件，则称其为中期个体理性（中期 *IR*）

$$U_i(t_i; M, \sigma) \geq o_i(t_i) \qquad \text{for every } i \text{ and } t_i \in T_i. \tag{12}$$

令 $\mathbf{M}(E)$ 表示此拟线性环境中的有限机制，令 $\mathbf{M}^I(E)$

包含那些机制，其每个均衡均满足(12)，并令 $\mathcal{D}^{T}(E)$ 为其真实直接成员。写出

$$\Pi^{*}(E) := \sup\left\{\Pi(M,\sigma)\colon\, M \in \mathcal{M}(E),\ \sigma \in \mathsf{Eq}(M) \text{ satisfies } (12)\right\}.$$

命题 6.1。 对于每个有限机制 M、每个中期个体理性均衡 $\sigma \in$方程(M)以及每个 $\varepsilon > 0$，都存在一个真实直接机制 D_{ε}，其唯一的贝叶斯纳什均衡是中期个体理性的，并满足 $\left|\Pi(D_{\varepsilon}、\ \sigma^{T}) - \Pi(M、\ \sigma)\right| < \varepsilon$。

其要点是，对于每个内部 V，

$$\sup_{D\in\mathcal{D}^{T}(E)} \mathcal{V}(S(D)) = \sup_{M\in\mathcal{M}^{I}(E)} \mathcal{V}(S(M)) = \Pi^{*}(E).$$

Others[17] 通过扩大代理人的消息空间获得其实施结果，而 Ollár 和 Penta (2017年) 则在具有特定光滑结构的直接机制中实现完全实施。[18] 我从任意具有有限类型空间、独立类型和私人价值的严格贝叶斯激励相容直接机制中生成唯一的真实贝叶斯纳什均衡。

引理 6.2。 对于任意严格贝叶斯激励相容直接机制 $(q、\ \tau)$，存在一个支付规则$\hat{}\tau$，使得 $(q,\ \hat{}\tau)$ 以如实报告作为其唯一的贝叶斯纳什均衡，并且

$$\sum_{t_{-i}} p_{-i}(t_{-i})\hat{\tau}_i(r_i,t_{-i}) = \sum_{t_{-i}} p_{-i}(t_{-i})\tau_i(r_i,t_{-i}) \qquad \textit{for every } i \textit{ and } r_i \in T_i. \tag{13}$$

我将在命题 6.1的附录 A.7中证明这个引理。

A 省略的证明

A.1 引理 4.1 的证明

引理 *4.1* 的证明引理 *4.1*。 考虑以下环境和机制。令 $N :=k+1$ 以及 $T_1 = T_2 = \{0,\ .....,\ k\}$。代理人的类型相互独立，且每种类型均

[17]例如，Abreu 和 Matsushima (1992)、d'Aspremont 等人(2005年)以及Kim (2013)。

[18]可微的估值、配置和转移函数；每个代理人对自身报告的期望收益严格凹；以及对手的报告影响激励的程度有界。

均匀分布于{0，……，k}上。令 C为类型集合之外的一个符号，并令 $X := (T_1 \cup \{C\}) \times (T_2 \cup \{C\})$。用 $x = (x_1, x_2)$ 表示一个结果。当代理人报告为 (r_1, r_2) 时，直接机制 D^0 以概率 1 选择结果 $x = (r_1, r_2)$。

对于 $x_i \in T_i$，令 j_i 为 $\{0,\ldots,k\}$ 中与 $x_i - t_i$ 模 N 同余的唯一整数。若 $x_i = C$，令 $j_i = C$，使得 $j_i = 0$ 当且仅当 $x_i = t_i$。在 D^0 下，类型 t_i 可通过报告 $(t_i + j)$ 模 N 来选择任意差值 $j_i = j$。

对于 $x_1 \in T_1$ 和 $x_2 \in T_2$，给予代理人 i 效用 $u_i(x,t) = -(j_i - j_{-i})^2$。若 $x_1 = C$ 或 $x_2 = C$，则将代理人 i 的效用定义为 $u_i(x,t) = \tilde{u}_i(j_i, j_{-i})$，其中

$$\tilde{u}_i(C,0) = -\frac{1}{2} = \tilde{u}_i(C,C), \quad \tilde{u}_i(0,C) = 0, \quad \& \quad -\tilde{u}_i(j,C) = \tilde{u}_i(C,j) = 1 \quad (j = 1,\ldots,k).$$

我选择委托人的支付，使具有共同差值的均衡取得 F 中的价值，同时混合均衡不会引入额外的支付。为此，定义对称矩阵 W 如下：

$$W_{jj} = a_j \quad (j = 0,\ldots,k), \quad \& \quad W_{j,j+1} = W_{j+1,j} = \frac{3a_{j+1} - a_j}{2} \quad (j = 0,\ldots,k-1), \tag{14}$$

并将所有其他非对角元素设为 a_k。若对于两个代理人均有 $x_i \in T_i$，令 $v(x,t) = W_{j_1 j_2}$；否则，令 $v(x,t) = a_k$。

只有 W_{00} 等于 b。所有其他对角元素均低于 b，且相邻元素满足 $(3a_{j+1} - a_j)/2 < a_{j+1} < b$。因此，委托人的收益从不超过 b，并且在类型组合 t 下达到 b 的唯一结果是 $x = t$。不同的类型组合对应不同的最大化结果。这证明了 (i) 部分。

断言 A.1。直接机制具有 $S(D^0) = F$，并且 F 中的每个成员都在严格纯策略均衡中实现。

证明。</style}在 D^0 下固定一个策略组合。令 $q_i(j)$ 表示无条件概率

代理人 i的差值等于 j:

$$q_i(j) \coloneqq \frac{1}{N}\sum_{t_i=0}^{k} \sigma_i\left((t_i + j) \bmod N \mid t_i\right) \qquad (j = 0, \ldots, k).$$

类型独立性意味着，代理人的每一种类型 i 都面临相同的q_{-i} 另一代理人的差异分布 q_i。由于每种类型都可以选择每一种差异，在均衡中，被 q_{-i}赋予严格正概率的每一种差异都必须最大化期望效用。逆命题是，在两个代理人选择差异并获得效用 $-(j_i - j_{-i})^2$ 的有限博弈的任何均衡中，都可以诱导出D^0：每种类型都采用该均衡下关于差异的分布。

面对差异均值为 μ的对手，选择差异 j 所获得的期望效用为 $-j^2 + 2j\mu - \mathbb{E}[j_{-i}^2]$。最佳回应是最接近 μ的整数。如果一个代理人选择纯差异 j，另一个代理人的唯一最佳回应是 j，因此双方都会选择该差异。如果两个分布都不是退化分布，则每个分布都必须恰好在两个相邻整数上赋予严格正概率。假设第一个代理人使用$\{j, j+1\}$，第二个代理人使用$\{\ell, \ell+1\}$。第一个代理人的无差异条件要求第二个代理人的均值为 $j + 1/2$；该均值严格位于 ℓ 与 $\ell + 1$之间，因此 $j = \ell$。随后，第二个代理人的无差异条件要求第一个代理人的均值也等于 $j + 1/2$。因此，双方都以概率 $1/2$ 分别赋予 j 和 $j + 1$。

这列举了均衡所诱导的所有可能的对（q_1， q_2），其中包括不同类型采用不同混合策略的均衡。在共同的纯差异 j下，委托人获得 a_j。在 j 和 $j + 1$上的共同等概率混合下，由于类型和私人随机化彼此独立，这两个差异也是相互独立的。因此，委托人的期望收益为

$$\frac{1}{4}a_j + \frac{1}{4}a_{j+1} + \frac{1}{2}\frac{3a_{j+1} - a_j}{2} = a_{j+1}. \tag{15}$$

因此，每个均衡收益都属于 F。

对于每个 $j \in \{0, \ldots, k\}$，让两个代理人的每种类型都报告 $(t_i + j) \bmod N$。每种类型获得的效用为 0，而任何不同的报告所带来的效用至多为 -1。

这是一个严格纯策略均衡，委托人获得 a_j。这些均衡涵盖了 F 中的每个元素，从而证明了该结论。 $j = 0$ 的情形是如实报告，因此 D^0 是真实的。■

断言 A.2。 机制 M^+为每个代理人 i 提供消息集$T_i \cup \{C\}$，并在消息组合 a之后以概率 *1* 选择 $x = a$ ，该机制具有唯一均衡，在该均衡中两位代理人都如实报告。

证明。 该博弈是支配可解的：首先，对于每个代理人和每种类型， C 严格支配虚假报告。其次，在删除虚假报告后，如实报告严格支配 C。■

断言 A.3。 任何具有使委托人获得 b的均衡的直接机制，都是 D^0的重新标记。

证明。 这直接源于以下事实：如果委托人获得 b，对于每个类型组合 t，以严格正概率使用的每个报告组合都必须以概率 1 导向唯一的最大化结果 $x = t$ 。■

断言 A.1 至 A.3 分别证明了第 (ii)、(iii) 和 (iv) 部分。 ■

A.2 引理 4.2 的证明

引理 4.2 的证明。假设对于某个 $\eta > 0$，结论不成立。那么存在满足 max $S(D_m)$ $> b - \frac{1}{m}$且 $d_H(S(D_m)\ F) \geq \eta$ 的直接机制 D_m。对于每个 m，选择一个达到 max $S(D_m)$ 的均衡 σ_m 。直接规则的空间是有限乘积 $\Delta(X)^T$，策略组合的空间也是单纯形的有限乘积。两者都是紧致的。取一个子序列后，记 $D_m \to D^*$ 和 $\sigma_m \to \sigma^*$。

均衡不等式（1）关于结果规则和策略是连续的，因此在极限处仍然成立，所以 $\sigma^* \in$ 方程(D^*)。由于 $(D,\ \sigma) \mapsto \Pi(D,\ \sigma)$ 是连续的，并且$b - \frac{1}{m} < \Pi(D_m,$ $\sigma_m) \leq b,\ \ \Pi(D^*,\ \sigma^*) = b$。引理 4.1 的第 (iv) 部分意味着 D^*

是 D^0 的重新标记，因此 $S(D^*) = F$。 F 中的每个收益都在 D^* 的某个严格纯策略均衡处取得。

命题 A.4。 $d_H(S(D_m)\ F) \to 0$。

证明。 固定 $a \in F$ 以及达到 a的 D^* 严格纯策略均衡。当结果规则充分接近 D^*时，表示每种类型都偏好其指定消息的有限个严格不等式仍然成立。因此，对于所有充分大的 m，相同的纯策略组合都是 D_m 的均衡。其委托人收益收敛于a。由于 F 是有限的，这就证明了最大值$_{a\in F}$ dist(a， $S(D_m)$) $\to 0$。

反之，假设存在 D_m 的均衡收益，在某个子序列上始终至少距离 F $\varepsilon > 0$ 。选取达到该收益的 $\tau_m \in$ Eq(D_m)。再取一个子序列后，记 $\tau_m \to \tau^*$。对（1）取极限得到 $\tau^* \in$Eq(D^*)。由于（D， τ） $\mapsto \Pi$（D， τ）是连续的， Π（D_m， τ_m） $\to \Pi$（D^*， τ^*） $\in F$，矛盾。因此，最大值$_{s\in S(D_m)}$ dist(s， F) $\to 0$。■

命题 A.4 与 d_H（S（D_m） F） $\geq \eta$相矛盾，证明完毕。 ■

A.3 引理 5.1 的证明

引理 5.1 的证明。固定一个机制 M 以及一个全支持先验序列 $p_m \to p$。对于每个 m，选择达到最大值 $S_{p_m}(M)$ 的 $\sigma_m \in$ Eq$_{p_m}(M)$。沿着实现上极限的一个子序列，策略空间的紧致性意味着 $\sigma_m \to \sigma$。均衡不等式关于全支持单纯形上的先验和策略组合连续，因此 $\sigma \in$ Eq$_p(M)$。委托人的期望收益也连续。因此，lim sup$_m$ max $S_{p_m}(M) = \Pi_p(M, \sigma) \leq$ max $S_p(M)$。■

A.4 引理 5.2 的证明

引理*5.2*的证明。 固定一个 i（至少有两种类型）、一种类型 t_i以及一个可能的报告 r_i。 在直接机制中要求其他代理人作如实报告时，报告 t_i 和 r_i 所得到的期望效用之差可写为

$\sum_{t_{-i}} p(t_{-i} \mid t_i)\Delta_i(t_{-i}; t_i,\ r_i)$，其中 Δ_i 由 M 的结果规则以及 σ^*中固定的消息分布决定。在 p^*下，该表达式非负，因为 σ^*是 M的均衡。共同中期信念使得显示式中的条件概率在每个 $p \in$ P下都相同。因此，同一个如实报告不等式在每个可能的先验分布下都成立。只有一种类型的代理人没有其他可供选择的直接报告。因此，该直接机制是稳健真实的。■

A.5 引理 5.3 的证明

引理 5.3 的证明。共有两位代理人。代理人 1 的类型集为单元素集 $T_1 =\{*\}$，代理人 2 的类型集为 $T_2 = \{0, 1, 2\}$。选择 $L <$最小值 F和 $H > 2b - L$。对于每个 $j = 1, \ldots, k$，令 $\beta := (b - L)/(H - L)$ 和 $\alpha_j := (a_j - L)/(H - L)$，并定义$p_j^0 := (\beta, \alpha_j, 1 - \beta - \alpha_j)$ 和 $p_j^1 := (\alpha_j, \beta, 1 - \beta - \alpha_j)$。令 $\mathrm{P} := \{p_j^0, p_j^1 :\ j = 1, \ldots, k\}$。唯一拥有多个类型的代理人，在任何先验分布下对代理人 1 都持有相同的退化信念，因此 P 具有共同中间信念。

共有两个结果，$X = \{0,1\}$。在代理人 2 的类型分别为 0、1 和 2 时，代理人 1 和委托人分别获得结果—收益对 (H, L)、(L, H) 和 (L, L)。代理人 2 从零结果获得效用 1，从结果 1 获得效用 0。

考虑间接机制 M^*。代理人 1 从$\{0, 1\}$中选择一条消息，代理人 2 只有一条消息，结果等于代理人 1 的消息。在 p_j^0 下，结果 0 和 1 分别给代理人 1 和委托人带来期望收益 b 和 a_j。在 p_j^1 下，它们带来的期望收益分别为 a_j 和 b。该机制在任意先验分布下都有唯一均衡，并为委托人带来 b。因此，$C_{\mathscr{P}}(M^*) = \{b\}$ 且 $\mathrm{A}(C_{\mathscr{P}}(M^*)) = b$。

在先验分布 P 中的任何先验下，都没有任何机制能为委托人带来高于 b 的均衡收益，因为代理人 2 的每种类型对结果都有相同的偏好，因此代理人 2 作出报告后，零结果的共同概率必定为 q。因此，对于稳健真实的直接机制，委托人在 p_j^0 和 p_j^1 下的均衡收益分别为 $qb + (1 - q)a_j$ 和 $qa_j + (1 - q)b$。它们的均衡收益

集合都是单点集，因此 (11) 也是先验特定值的集合。每个 $q \in [0,1]$ 都可以通过在每份报告后以 q 的概率选择零结果来实现。■

A.6 引理 5.4 的证明

引理 *5.4* 的证明。 由于共同中期信念不成立，存在 p、$q \in \mathcal{P}$、一个 i 至少具有两种类型的代理人，以及一个类型 t_i^0，使得 $p(\cdot \mid t_i^0) \neq q(\cdot \mid t_i^0)$。选择一个函数 $d\colon T_{-i} \to \mathbf{R}$ 满足

$$\sum_{t_{-i}} p(t_{-i} \mid t_i^0) d(t_{-i}) > 0 > \sum_{t_{-i}} q(t_{-i} \mid t_i^0) d(t_{-i}). \tag{16}$$

由于两个条件分布不同，这样的函数确实存在：选择一个它们赋予不同概率的事件，并减去一个严格处于这两个概率之间的常数。

取 $X = T$，并使委托人收益为 $v(x, t) = \mathbf{1}_{\{x=t\}}$。对于每个代理人 $j \neq i$，设 $u_j(x, t_j) = \mathbf{1}_{\{x_j=t_j\}}$，并对于代理人 i，设

$$u_i(x, t_i) = \begin{cases} d(x_{-i}), & t_i = t_i^0 \text{ and } x_i = t_i^0, \\ 0, & t_i = t_i^0 \text{ and } x_i \neq t_i^0, \\ \mathbf{1}_{\{x_i=t_i\}}, & t_i \neq t_i^0. \end{cases}$$

令 D^0 在代理人报告 $r \in T$时必然选择结果 $x = r$。在先验分布 p下，如实报告是严格均衡，因为代理人 i 的类型 t_i^0根据（16）中的第一个不等式严格偏好如实报告，而其他每一种类型如实报告时均获得效用一，作出任何虚假报告时均获得零。委托人获得一，任何机制都不能超过上界一，因此无限制价值为一。

考虑一个直接机制 D，在某个全支持先验分布下，该机制存在一个使委托人获得 1 的均衡。对于每个类型组合 t，D 在代理人以严格正概率采用的每个报告组合之后，都必须以概率 1 指派 $x = t$。根据断言 A.3的证明中的论证，每种类型都会使用一个报告，且同一代理人的不同类型

会使用不同的报告。令每个代理人 j 按照 π_j报告：在该均衡中为 $T_j \to T_j$。每个 π_j 都是一个双射；在报告 r、 D 之后， $x = \left(\pi_1^{-1}$ 以概率 1 指派结果 r_1(.....，$\pi_n^{-1}(r_n)\right)$。

假设 D 在 p 和 q下都是真实的。每个代理人的每一种类型 t_j ，除了代理人的类型 t_i^0 以外， j都从报告 $\pi_j(t_j)$ 中获得效用 1，而从任何其他报告中获得效用 0，无论其他代理人的报告是什么。因此，对于所有这些类型，真实性都要求 $\pi_j(t_j) = t_j$。由于 π_i 是一个双射，并且将其他每一种类型映射到自身，因此它也必须将 t_i^0 映射到自身。因此， $D = D^0$。然而，在 q下，当其他代理人如实报告时，类型 t_i^0 如实报告会获得严格为负的期望效用，这是（16）中的第二个不等式所表明的。任何虚假报告都会给他带来零收益，因此 D 不可能在两个先验分布下都是真实的。

令 Γ 为满足以下条件的三元组（D， r， σ）： D 是在 p 和 q、 $r \in$P下都真实的直接规则，并且 $\sigma \in$Eq$_r$（D）。该集合是紧致的。[19] 中的任何三元组都无法使委托人收益达到 1。委托人的期望收益 Π_r（D， σ）关于（D， r， σ）是连续的，因此它在 Γ上取得最大值 $\bar{v} < 1$。每个稳健真实直接机制都在 p 和 q下是真实的，因此其在任意 $r \in$P下的所有均衡收益都不超过 $\bar{v}$，所以 W^R（E^0，P） $\leq \bar{v}$。■

A.7 命题 6.1 与 引理 6.2的证明

引理 *6.2* 的证明。如果至多有一位代理人的类型数超过一个，那么严格激励相容性已经使如实报告成为唯一均衡，因此令 $\hat{\tau} = \tau$。否则，对于每位代理人 i，固定一个任意报告 $t_i^0 \in T_i$。对于每个 $a \in T_i \setminus \{t_i^0\}$，定义 $\phi_{ia}(r_i) := \mathbf{1}_{\{r_i=a\}} - p_i(a)$。

现在，我将代理人 i 的支付从 $\tau_i(r)$ 修改为 $\tau_i(r) - \kappa_i H_i(r)$，具体如下。我选择$H_i$，使得对于每个报告 r_i，当其对手如实报告时，代理人 i 的期望支付在 $\tau_i - \kappa_i H_i$ 下与在 τ_i 下相同。由于原机制是严格贝叶斯激励相容的，当其对手如实报告时，他会严格倾向于如实报告。关键在于，

[19]直接规则的空间和策略空间都是紧致的，在 p 和 q 下的真实性由有限个闭不等式给出，P 是紧致的，并且均衡不等式在 (D, r, σ) 中具有闭图。

这些修改可能会影响对手未如实报告时对他的激励，我将利用这一点消除其他均衡。

现在来构造这些对象。首先，对每个代理人和报告组合，令 $H_i(r) = 0$。对于每个代理人 i，考虑 $T_i \setminus \{t_i^0\}$ 中的报告。选择两个代理人 $i < j$，并分别从他们的报告中选取一个，例如 a 和 b。将 $\phi_{ia}(r_i)\phi_{jb}(r_j)$ 加到 $H_i(r)$ 上，并从 $H_j(r)$ 中减去。重复此过程，但不重复选择同一代理人的同一报告，直到无法再选出报告对为止。

如果在此构造中还有未使用的报告，那么它们必定全都属于同一位代理人，否则我就可以再选出一对报告。将该代理人记为 i_0。如果没有剩余报告，则任选一位类型数多于一个的代理人作为 i_0。每个乘积都被加到一个函数上，并从另一个函数中减去，因此对每个 r，$\sum_i H_i(r) = 0$。

$H_i(r)$ 中的每一项，对于某个 $j \neq i$，要么是 $\phi_{ia}(r_i)\phi_{jb}(r_j)$，要么是其相反数。固定 r_i，并假设对手如实报告。于是：

$$\sum_{t_{-i}} p_{-i}(t_{-i})\phi_{ia}(r_i)\phi_{jb}(t_j) = \phi_{ia}(r_i)\sum_{t_j} p_j(t_j)\left[\mathbf{1}_{\{t_j=b\}} - p_j(b)\right] = 0.$$

反过来，固定 r_{-i}，并假设代理人 i 如实报告，在这种情况下

$$\sum_{t_i} p_i(t_i)\phi_{ia}(t_i)\phi_{jb}(r_j) = \phi_{jb}(r_j)\sum_{t_i} p_i(t_i)\left[\mathbf{1}_{\{t_i=a\}} - p_i(a)\right] = 0.$$

因此，

$$\sum_{t_{-i}} p_{-i}(t_{-i})H_i(r_i, t_{-i}) = 0 = \sum_{t_i} p_i(t_i)H_i(t_i, r_{-i}) \tag{17}$$

对于每个代理人 i、每个报告 $r_i \in T_i$，以及对手报告的每个策略组合 $r_{-i} \in T_{-i}$，均成立。

现在考虑一个没有私人信息的辅助博弈。每个代理人 i只需选择 $r_i \in T_i$，并在由此形成的策略组合上获得 $H_i(r)$。

命题 A.5。辅助博弈的每个纳什均衡 $\mu = (\mu_i)_i$ 都满足：对所有 $i \neq i_0$，$\mu_i = p_i$。

证明。固定 辅助博弈的一个纳什均衡 μ。根据

式（17）中的第二个等式，$\sum_{r_i} p_i(r_i) \sum_{r_{-i}} \left(\prod_{j\neq i} \mu_j(r_j)\right) H_i(r_i, r_{-i}) = 0$，因此代理人 i 的均衡收益非负。此外，$\sum_i \sum_r \left(\prod_j \mu_j(r_j)\right) H_i(r) = \sum_r \left(\prod_j \mu_j(r_j)\right) \sum_i H_i(r) = 0$，因此每位代理人的均衡收益都为零。因此，对于每个 $r_i \in T_i$，

$$\sum_{r_{-i}} \left(\prod_{j\neq i} \mu_j(r_j)\right) H_i(r_i, r_{-i}) \leq 0. \tag{18}$$

但 $\sum_{r_i} p_i(r_i) \sum_{r_{-i}} \left(\prod_{j\neq i} \mu_j(r_j)\right) H_i(r_i, r_{-i}) = 0$，并且由于对每个 r_i，$p_i(r_i) > 0$ 都成立，因此对每个 $r_i \in T_i$ 都必须取等号。 (18)

取在上述构造中一起选出的报告 $a \in T_i \setminus \{t_i^0\}$ 和 $b \in T_j \setminus \{t_j^0\}$，其中 $i < j$。该构造将 $\phi_{ia}(r_i)\phi_{jb}(r_j)$ 加到 $H_i(r)$ 上，并从 $H_j(r)$ 中减去它。由于没有任何代理人的报告被重复选取，这是 H_i 中唯一包含 ϕ_{ia} 的项，也是 H_j 中唯一包含 ϕ_{jb} 的项。此外，$\phi_{ia}(a) - \phi_{ia}(t_i^0) = 1$，而对每个 $c \in T_i \setminus \{t_i^0, a\}$，$\phi_{ic}(a) - \phi_{ic}(t_i^0) = 0$ 都成立。由于 a 和 t_i^0 在面对 μ_{-i} 时都使代理人 i 的期望收益为零，

$$0 = \sum_{r_{-i}} \left(\prod_{k\neq i} \mu_k(r_k)\right) \left[H_i(a, r_{-i}) - H_i(t_i^0, r_{-i})\right] = \sum_{r_j} \mu_j(r_j)\phi_{jb}(r_j) = \mu_j(b) - p_j(b).$$

因此，$\mu_j(b) = p_j(b)$。以同样的逻辑，$\mu_i(a) = p_i(a)$。

我们已经推导出，在每一对中所选取的两个报告处，μ_i 与 p_i 相等。根据 i_0 的选择，每个代理人 $i \neq i_0$ 的 $T_i \setminus \{t_i^0\}$ 中的每份报告都在其中一对中被选取。因此，对于每个 $i \neq i_0$ 以及每个 $a \in T_i \setminus \{t_i^0\}$，$\mu_i(a) = p_i(a)$。由于 μ_i 和 p_i 是分布，$\mu_i(t_i^0) = 1 - \sum_{a\in T_i\setminus\{t_i^0\}} \mu_i(a) = 1 - \sum_{a\in T_i\setminus\{t_i^0\}} p_i(a) = p_i(t_i^0)$。■

对于 $K \geq 1$，令 $\kappa_{i_0}(K) = K$ 且 $\kappa_i(K) = K^2$ 对于 $i \neq i_0$ 成立，并将支付修改为

$$\hat{\tau}_i^K(r) := \tau_i(r) - \kappa_i(K) H_i(r). \tag{19}$$

由（19）以及（17）中的第一个等式可得，

$$\sum_{t_{-i}} p_{-i}(t_{-i})\hat{\tau}_i^K(r_i, t_{-i}) = \sum_{t_{-i}} p_{-i}(t_{-i})\tau_i(r_i, t_{-i}) - \kappa_i(K) \sum_{t_{-i}} p_{-i}(t_{-i}) H_i(r_i, t_{-i}) = \sum_{t_{-i}} p_{-i}(t_{-i})\tau_i(r_i, t_{-i}),$$

这证明了（13）。当对手如实报告时，它也使报告之间的每一项事前比较保持不

变。因此，对任意 K，如实报告仍是严格的贝叶斯纳什均衡。

我证明，对于所有充分大的有限 K，真实报告是唯一的均衡。反设不然。那么存在一个序列 $K_m \to \infty$，记 (19) 中在 $K = K_m$ 时的支付规则为 $\hat{\tau}^m$，则对每个 m，$(q, \hat{\tau}^m)$ 都存在一个非真实均衡 σ^m。定义代理人 i 报告 r_i 的事前概率为 $\mu_i^m(r_i) := \sum_{t_i} p_i(t_i)\sigma_i^m(r_i \mid t_i)$，并定义$\mu^m := (\mu_i^m)_i \in \prod_i \Delta(T_i)$，表示由此得到的报告分布组合。由于代理人的类型相互独立，报告组合 r 的概率为

$$\sum_{t\in T}\prod_i p_i(t_i)\sigma_i^m(r_i \mid t_i) = \prod_i \sum_{t_i} p_i(t_i)\sigma_i^m(r_i \mid t_i) = \prod_i \mu_i^m(r_i).$$

如果代理人 i 的类型为 t_i，并报告 r_i，那么在原机制 (q, τ) 下，他的期望净效用为 $\sum_{r_{-i}} \left(\prod_{j\neq i} \mu_j^m(r_j)\right) [u_i(q(r_i, r_{-i}), t_i) - \tau_i(r_i, r_{-i})]$。

对每个代理人、类型和报告组合，选择 $B > 0$ 使得 $|u_i(q(r), t_i) - \tau_i(r)| \leq B$。[20] 取一个子序列，记作 $\mu^m \to \mu$。对于报告分布的策略组合 $\nu = (\nu_j)_j \in \prod_j \Delta(T_j)$，定义

$$H_i(\nu) := \sum_{r\in T}\left(\prod_j \nu_j(r_j)\right) H_i(r) \quad \& \quad H_i(r_i, \nu_{-i}) := \sum_{r_{-i}\in T_{-i}}\left(\prod_{j\neq i} \nu_j(r_j)\right) H_i(r_i, r_{-i}).$$

固定一个代理人 i 和一个报告 $r_i \in T_i$。由于 σ^m 是 $(q, \hat{\tau}^m)$ 的贝叶斯纳什均衡，每种类型 $t_i \in T_i$从遵循 $\sigma_i^m(\cdot \mid t_i)$ 中获得的期望效用至少与报告r_i时一样高，因此

$$\begin{aligned}
&\sum_{s_i, r_{-i}} \sigma_i^m(s_i \mid t_i)\left(\prod_{j\neq i} \mu_j^m(r_j)\right) [u_i(q(s_i, r_{-i}), t_i) - \tau_i(s_i, r_{-i}) + \kappa_i(K_m)H_i(s_i, r_{-i})] \\
&\qquad \geq \sum_{r_{-i}}\left(\prod_{j\neq i} \mu_j^m(r_j)\right) [u_i(q(r_i, r_{-i}), t_i) - \tau_i(r_i, r_{-i}) + \kappa_i(K_m)H_i(r_i, r_{-i})] .
\end{aligned}$$

将这些不等式分别乘以 $p_i(t_i)$，并对 t_i 求和。包含 H_i 的各项变为

[20]由于原机制固定，且其类型集合和报告集合均有限，因此这样的界存在。

$\kappa_i(K_m)H_i(\mu^m)$ 和 $\kappa_i(K_m)H_i(r_i,\ \mu^m_{-i})$。根据 B 的选择，

$$-B \le \sum_{t_i,s_i,r_{-i}} p_i(t_i)\sigma_i^m(s_i \mid t_i)\left(\prod_{j\neq i}\mu_j^m(r_j)\right)[u_i(q(s_i,r_{-i}),t_i)-\tau_i(s_i,r_{-i})] \le B, \quad \&$$

$$-B \le \sum_{t_i,r_{-i}} p_i(t_i)\left(\prod_{j\neq i}\mu_j^m(r_j)\right)[u_i(q(r_i,r_{-i}),t_i)-\tau_i(r_i,r_{-i})] \le B.$$

因此，$B + \kappa_i(K_m)H_i(\mu^m) \ge -B + \kappa_i(K_m)H_i(r_i,\mu^m_{-i})$，所以 $H_i(\mu^m) \ge H_i(r_i, \mu^m_{-i}) - \frac{2B}{\kappa_i(K_m)}$。

由于 $m\to\infty$, $\kappa_i(K_m)\to\infty$。$H_i(\nu)$ 和 $H_i(r_i,\nu_{-i})$ 都关于 ν 连续，因此 $\mu^m\to\mu$ 意味着 $H_i(\mu^m)\to H_i(\mu)$ 且 $H_i(r_i,\mu^m_{-i})\to H_i(r_i,\mu_{-i})$。将 $m\to\infty$ 代入 $H_i(\mu^m)\ge H_i(r_i,\mu^m_{-i}) - \frac{2B}{\kappa_i(K_m)}$，得到对于每个 i 和 r_i，均有 $H_i(\mu)\ge H_i(r_i,\mu_{-i})$。因此，$\mu$是辅助博弈的纳什均衡。根据命题 A.5，$\mu_i = p_i$ 对每个 $i\neq i_0$都成立。由于 $\mu^m\to\mu$，对每个 $i\neq i_0$，都有 $\mu_i^m\to p_i$。

令 $e_i^m := \sum_{t_i} p_i(t_i)[1-\sigma_i^m(t_i\mid t_i)]$ 为代理人 i的误报概率。定义

$$d_i^m := \sum_{t_i,r_i,r_{-i}} p_i(t_i)\sigma_i^m(r_i\mid t_i)\left(\prod_{j\neq i}\mu_j^m(r_j)\right) \times [u_i(q(t_i,r_{-i}),t_i)-\tau_i(t_i,r_{-i})-u_i(q(r_i,r_{-i}),t_i)+\tau_i(r_i,r_{-i})].$$

每个系数 $p_i(t_i)\sigma_i^m(r_i\mid t_i)\left(\prod_{j\neq i}\mu_j^m(r_j)\right)$ 均为非负数，并且对t_i、r_i和 r_{-i}求和为 1。每个方括号内的表达式都介于 $-2B$ 和 $2B$之间，因此 $|d_i^m|\le 2B$。

对于拥有多个类型的每个代理人 i，严格激励相容性为

$$\sum_{r_{-i}}\left(\prod_{j\neq i}p_j(r_j)\right)[u_i(q(t_i,r_{-i}),t_i)-\tau_i(t_i,r_{-i})-u_i(q(r_i,r_{-i}),t_i)+\tau_i(r_i,r_{-i})] > 0$$

对于每个 $t_i\in T_i$ 和每个 $r_i\neq t_i$。这些表达式关于对手的报告分布是连续的。由于代理人、类型和报告的数量都是有限的，因此存在 $c>0$，并且对于每个具有多个类型的代理人 i，存在 $(p_j)_{j\neq i}$的一个邻域 N_i，使得

$$(\nu_j)_{j\neq i}\in N_i \implies \sum_{r_{-i}}\left(\prod_{j\neq i}\nu_j(r_j)\right)[u_i(q(t_i,r_{-i}),t_i)-\tau_i(t_i,r_{-i})-u_i(q(r_i,r_{-i}),t_i)+\tau_i(r_i,r_{-i})]\ge c$$

对于每个 $t_i \in T_i$ 和每个 $r_i \neq t_i$。因此，

$$(\mu_j^m)_{j\neq i} \in N_i \quad \Longrightarrow \quad d_i^m \geq c\sum_{t_i} p_i(t_i) \sum_{r_i \neq t_i} \sigma_i^m(r_i \mid t_i) = ce_i^m. \tag{20}$$

若 $T_i = \{t_i\}$，则 $\sigma_i^m(t_i \mid t_i) = 1$，因此 $e_i^m = d_i^m = 0$。对于该代理人，令 $N_i = \prod_{j\neq i} \Delta(T_j)$。

若代理人 i 的每种类型都如实报告，而其他代理人采用 σ_{-i}^m，则代理人i的类型服从 p_i分布，而对手的报告服从 $\prod_{j\neq i} \mu_j^m$分布。因此，根据式 (17) 中的第二个等式，$\sum_{r_{-i}} \left(\prod_{j\neq i} \mu_j^m(r_j)\right) \sum_{t_i} p_i(t_i) H_i(t_i, r_{-i}) = 0$。于是，在 $(q, \hat{\tau}^m)$ 下，代理人 i 以如实报告替换 σ_i^m后，其事前效用的变化为 $d_i^m - \kappa_i(K_m) H_i(\mu^m)$；由于 σ_i^m是对 σ_{-i}^m的最佳响应，该变化非正。因此，$d_i^m \leq \kappa_i(K_m) H_i(\mu^m)$，并且

$$\sum_i \frac{d_i^m}{\kappa_i(K_m)} \leq \sum_i H_i(\mu^m) = \sum_{r\in T} \left(\prod_j \mu_j^m(r_j)\right) \sum_i H_i(r) = 0, \quad \text{so} \quad K_m d_{i_0}^m + \sum_{i\neq i_0} d_i^m \leq 0. \tag{21}$$

由于对于每个 $\mu_j^m \to p_j$，$j \neq i_0$，$(\mu_j^m)_{j\neq i_0} \in N_{i_0}$ 对所有充分大的 m 都成立。对于这些 m，

$$0 \leq ce_{i_0}^m \leq d_{i_0}^m \leq -\frac{1}{K_m} \sum_{i\neq i_0} d_i^m \leq \frac{1}{K_m} \sum_{i\neq i_0} |d_i^m| \leq \frac{2B(n-1)}{K_m} \to 0.$$

如实报告的分布为 p_{i_0}，因此 $\|\mu_{i_0}^m - p_{i_0}\|_1 \leq 2e_{i_0}^m \to 0$。因此，所有代理人的报告分布都收敛于其真实分布。

我们已经证明，对于每个代理人 j，$\mu_j^m \to p_j$。因此，对于所有充分大的m，对于每个代理人 i，$(\mu_j^m)_{j\neq i} \in N_i$，并且对于每个 i，$d_i^m \geq ce_i^m \geq 0$。同时，(21)意味着 $\sum_i \frac{d_i^m}{\kappa_i(K_m)} \leq 0$。该和式中的每一项都非负，因此对于每个 i，$d_i^m = 0$。由于$0 = d_i^m \geq ce_i^m$且 $c > 0$，对于每个 i，$e_i^m = 0$。最后，$0 = e_i^m = \sum_{t_i} p_i(t_i)[1 - \sigma_i^m(t_i \mid t_i)]$。该和式中的每一项都非负，并且对于每种类型，$p_i(t_i) > 0$。因此，对于每个 i和 t_i，$\sigma_i^m(t_i \mid t_i) = 1$，这与 σ^m的选取相矛盾。因此，对于充分大的有限 K，(19)中的支付规则使如实报告成为唯一的贝叶斯纳什均衡。取任意一个这样的 K 即可完成证明。■

证明 命题 6.1。 固定一个有限机制 $M = ((A_i)_i,\ q,\ \tau)$ 以及一个σ 满足个体理性的均衡。通过令下式成立，将该均衡直接化：

$$q^0(x \mid r) \coloneqq \sum_{a \in A} q(x \mid a) \prod_i \sigma_i(a_i \mid r_i), \qquad \& \qquad \tau_i^0(r) \coloneqq \sum_{a \in A} \tau_i(a) \prod_j \sigma_j(a_j \mid r_j).$$

如果代理人如实报告，该直接机制 $D^0 = (q^0,\ \tau^0)$ 将重现条件配置概率以及每种类型的中期预期支付。类型为 t_i 的报告r_i 恰好诱导出在原机制中针对 σ_{-i} 使用$\sigma_i(\cdot \mid r_i)$ 所得到的配置彩票和期望支付。该混合消息对于 t_i 类型在 M 中是可用的，因此均衡不等式使得在 D^0 中如实报告成为最优选择。结果是，对于每个 i 和 t_i，$U_i(t_i;\ D^0,\ \sigma^T) = U_i(t_i;\ M,\ \sigma)$，且 $\Pi(D^0,\ \sigma^T) = \Pi(M,\ \sigma)$。

接下来，我构造一个使每个如实报告不等式都严格成立的直接机制。对于每个代理人和类型，定义

$$w_i(t_i)_x \coloneqq u_i(x, t_i) - \frac{1}{|X|} \sum_{y \in X} u_i(y, t_i) \qquad (x \in X).$$

根据构造，对于每个 i 和 t_i，$\sum_{x \in X} w_i(t_i)_x = 0$。根据假设，当 $t_i \neq s_i$ 时，$w_i(t_i) \neq w_i(s_i)$。选择一个在 X 上具有全支持的分配彩票 q°。由于代理人和类型的数量有限，存在某个 $\eta > 0$，使得对于每个i，r_i，$q^\circ + \eta w_i(r_i) \in \Delta(X)$。定义直接机制 $L = (q^L,\ \tau^L)$（范数取欧几里得范数）如下：

$$q^L(r) \coloneqq \frac{1}{n} \sum_i \left(q^\circ + \eta w_i(r_i)\right), \qquad \& \qquad \tau_i^L(r) \coloneqq \frac{\eta}{2n} \|w_i(r_i)\|^2.$$

对于对手报告的任意向量，类型 t_i从真实报告相较于另一份报告 r_i 中获得的净收益为

$$\begin{aligned} &u_i(q^L(t_i, r_{-i}), t_i) - \tau_i^L(t_i, r_{-i}) - u_i(q^L(r_i, r_{-i}), t_i) + \tau_i^L(r_i, r_{-i}) \\ &\quad = \frac{\eta}{n} \langle w_i(t_i), w_i(t_i) - w_i(r_i) \rangle - \frac{\eta}{2n} \left(\|w_i(t_i)\|^2 - \|w_i(r_i)\|^2 \right) \\ &\quad = \frac{\eta}{2n} \|w_i(t_i) - w_i(r_i)\|^2 > 0, \end{aligned}$$

其中，第一个等式利用了 $\sum_{x \in X} [w_i(t_i)_x - w_i(r_i)_x] = 0$。因此，如实报告

在 L 下是严格优势策略。

写出 $U_i^L(t_i) := U_i(t_i; L, \sigma^T)$，并令 $b_i := \max_{t_i \in T_i} \max\{0, o_i(t_i) - U_i^L(t_i)\}$。对于 $\delta \in (0, 1)$，定义 $q^\delta := (1-\delta)q^0 + \delta q^L$ 以及 $\tau_i^\delta := (1-\delta)\tau_i^0 + \delta\tau_i^L - \delta b_i - \delta^2$。由于 D^0 满足激励相容，L 使每个非平凡报告不等式都严格成立，并且每个 $-\delta b_i - \delta^2$ 都不取决于代理人 i 的报告，因此 $D^\delta = (q^\delta, \tau^\delta)$ 是严格贝叶斯激励相容的。

应用引理 6.2 于 D^δ，得到 $\hat{} D^\delta = (q^\delta, \hat{} \tau^\delta)$。如实报告是其唯一的贝叶斯纳什均衡。（13）保留了 D^δ 的真实中期支付和效用，因此该均衡仍满足中期个体理性。对于每种类型，

$$\sum_{t_{-i}} p_{-i}(t_{-i})\hat{\tau}_i^\delta(t_i, t_{-i}) = (1-\delta)\bar{\tau}_i^\sigma(t_i) + \delta \sum_{t_{-i}} p_{-i}(t_{-i})\tau_i^L(t_i, t_{-i}) - \delta b_i - \delta^2.$$

由于 $\delta \downarrow 0$，这些支付收敛于 $\bar{\tau}_i^\sigma(t_i)$，而对于每个 x，t，$q^\delta(x \mid t) \to q^0(x \mid t) = q^\sigma(x \mid t)$。因此，$\Pi(\hat{} D^\delta, \sigma^T) \to \Pi(M, \sigma)$。选择足够小的 $\delta > 0$，使得 $\left|\Pi(\widehat{D^\delta}, \sigma^T) - \Pi(M, \sigma)\right| < \varepsilon$，并令 $D_\varepsilon := \widehat{D^\delta}$。■

参考报告

Dilip Abreu 和 Hitoshi Matsushima。不完全信息下迭代剔除劣策略中的虚拟实施。*Mimeo*，1992年。

Dirk Bergemann 和 Stephen Morris。稳健机制设计。《计量经济学》，73（6）：1771–1813，2005年。

Dirk Bergemann 和 Stephen Morris。直接机制中的稳健实施。《经济学研究评论》，76（4）：1175–1204，2009年。

Dirk Bergemann 和 Stephen Morris。信息设计、贝叶斯劝说与贝叶斯相关均衡。《美国经济评论》，106（5）：586–591，2016年。

Dirk Bergemann 和 Stephen Morris。《信息设计：统一视角》。《经济文献杂志》，57(1):44–95，2019年。

Dirk Bergemann、Benjamin Brooks 和 Stephen Morris。信息稳健的最优拍卖设计。未发表手稿，2016年。

Dirk Bergemann、Stephen Morris 和 RafaelVeiel。《信息结构上的策略拓扑》。*Mimeo*，2024年。

Benjamin Brooks 和 Songzi Du。《共同价值下的最优拍卖设计：一种信息稳健方法》。《计量经济学》，89(3)：1313–1360，2021年。

Benjamin Brooks 和 Songzi Du。《信息稳健最优机制的结构》。《计量经济学》，92(5)：1391–1438，2024年。

Bernard Caillaud 和 Jacques Robert。《无信息卖方对收益最大化拍卖的实施》。《经济设计评论》，9(2)：127–143，2005年。

Jaewon Choi 和 Taesung Kim。《一种非参数、高效的公共物品决策机制：无支配贝叶斯实施》。《博弈与经济行为》，27(1)：64–85，1999年。

Kim-Sau Chung 和 Jeffrey C. Ely。《占优策略机制的基础》。《经济学研究评论》，74(2)：447–476，2007年。

Partha Dasgupta、Peter Hammond 和 Eric Maskin。社会选择规则的实施：关于激励相容性的一些一般性结果。《经济学研究评论》，46(2):185–216，1979年。

Claude d'Aspremont、Jacques Crémer 和 Louis-André Gérard-Varet。拍卖和公共品问题中的唯一实施。《公共经济学》*Économie publique/ Publiceconomics*，(17):125–139，2005年。

PeerapongDhangwatnotai、Tim Roughgarden 和 Qiqi Yan。利用单一样本实现收入最大化。《博弈与经济行为》，91：318–333，2015年。

Yiding Feng 和 Jason D. Hartline。机制设计中的端到端论证（面向预算受限代理人的先验无关拍卖）。载于*2018 IEEE 第59*届计算机科学基础年度研讨会（*FOCS*），第404–415页，2018年。

Yiding Feng、Jason D. Hartline 和 Yingkai Li。基于样本定价的揭示差距。载于第*53*届*ACM SIGACT*计算理论年会论文集，第1438–1451页。计算机协会，2021年。

Matthew O. Jackson。贝叶斯实施。《计量经济学》，59(2)：461–477，1991年。

Bong-JuKim。带侧支付的贝叶斯实施的一个充分条件。《韩国经济评论》，29(2)：429–445，2013年。

Ehud Lehrer、Eilon Solan 和 Yannick Viossat。有限博弈的均衡收益。《数学经济学杂志》，47(1)：48–53，2011年。

Laurent Mathevet。超模机制设计。《理论经济学》，5(3)：403–443，2010年。

Laurent Mathevet 和 Ina Taneva。具有相互依赖估值的有限超模设计。《博弈与经济行为》，82：327–349，2013年。

Laurent Mathevet、Jacopo Perego 和 Ina Taneva。论博弈中的信息设计。《政治经济学杂志》，128(4)：1370–1404，2024年。

Dilip Mookherjee 和 Stefan Reichelstein。通过扩展揭示机制实施。《经济学研究评论》，57(3)：453–475，1990。

Stephen Morris、Daisuke Oyama 和 Satoru Takahashi。通过二元行动超模博弈中的信息设计实施。《计量经济学》，92(3)：775–813，2024年。

Roger B. Myerson。《广义委托—代理问题中的最优协调机制》。《数学经济学杂志》，10(1):67–81，1982年。

Mariann Ollár和Antonio Penta。《完全实施与信念限制》。《美国经济评论》，107(8):2243–2277，2017年。

Mariann Ollár 和 Antonio Penta。稳健实施的网络解法：相同但未知分布的情形。《经济学研究评论》，90(5):2517–2554，2023年。

Thomas R. Palfrey 和 Sanjay Srivastava。交换经济中不完全信息下的实施。《计量经济学》，57(1):115–134，1989a。

Thomas R. Palfrey 和 Sanjay Srivastava。不完全信息下的机制设计：实施问题的一种解法。《政治经济学杂志》，97(3):668–691，1989b。

Rafael Repullo。完全和不完全信息下占优策略的实施。《经济学研究评论》，52(2):223–229，1985年。

Roberto Serrano 和 Rajiv Vohra。机制中混合均衡的多重性：精确实施与近似实施的统一方法。《数学经济学杂志》，46(5):775–785，2010年。

Ina Taneva。《信息设计》。《美国经济学杂志：微观经济学》，11 (4):151–185，2019年。

Mark Whitmeyer。《服从意味着什么》。未发表手稿，2026年。

B 补充：私人价值

现在，我证明定理 3.1 中关于私人价值和独立类型的论断。我修改了§4 中的构造，使代理人的效用只取决于他自己的类型和所选结果。类型仍然相互独立，但我用一种大部分概率都赋予类型零的分布取代均匀边际分布。

对于给定有限集 F 中的每一个收益，修改后的直接机制仍然具有一个使委托人获得该收益的严格均衡。尽管它可能还具有其他均衡收益，但通过将足够大的概率集中于类型零上，我可以使每个额外收益都任意接近 F。

引理 B.1。令 $F = \{a_0, a_1, \dots, a_k\}$，其中 $b := a_0 > a_1 > \cdots > a_k$ 以及 $k \geq 1$，并固定 $\eta > 0$。存在一个环境 E，其中有两个代理人、私人价值，以及定义在 $\{0, \dots, k\}$ 上的独立类型；还存在一个真实直接机制 D^0、一个为每个代理人增加一条额外消息的机制 M^+，以及一个 $\delta > 0$，满足：

(i) 在每个类型组合下，委托人的收益至多为 b，且恰有一个结果使她获得 b。这些最大化结果因类型组合而异。(ii) $F \subseteq S(D^0)$，$d_H(S(D^0)\ F) < \eta/2$，并且 F 中的每个成员都在 D^0 的某个严格纯策略均衡中实现。(iii) M^+ 具有唯一的贝叶斯纳什均衡，其中 $S(M^+) = \{b\}$。(iv) 每个直接机制 D 都满足

$$\max S(D) > b - \delta \quad \Longrightarrow \quad d_H(S(D), F) < \eta. \tag{22}$$

证明。令 $N := k+1$，令 $T_1 = T_2 = \{0, \dots, k\}$，并取 $X := (T_1 \cup \{C\}) \times (T_2 \cup \{C\})$，其中 C 是类型集合之外的一个符号。暂且固定 $0 < \varepsilon < 1/(2k^2 + 3)$。代理人的类型相互独立，服从共同分布 $p_\varepsilon(0) = 1 - \varepsilon$ 和 $p_\varepsilon(j) = \frac{\varepsilon}{k}$，其中 $j = 1, \dots, k$。在确立这族环境的性质后，我将选取 ε。

对于 $x_i \in T_i$，令 j_i 为 $\{0,\dots,k\}$ 中与 $x_i - t_i$ 模 N 同余的唯一整数。若 $x_1 \in T_1$ 且 $x_2 \in T_2$，则给予代理人 i 效用 $u_i(x,t_i) = -(j_i - x_{-i})^2$。若 $x_1 = C$ 或 $x_2 = C$，则设

$$u_i(x,t_i) = \begin{cases} -\frac{1}{2}, & x_i = C, \quad x_{-i} \in \{0, C\}, \\ 1, & x_i = C, \quad x_{-i} \in \{1,\dots,k\}, \\ 0, & x_i = t_i, \quad x_{-i} = C, \\ -1, & x_i \in T_i \setminus \{t_i\}, \quad x_{-i} = C. \end{cases}$$

保留引理 4.1 中委托人的收益。若 $x_1 \in T_1$ 且 $x_2 \in T_2$，令 $v(x,t) = W_{j_1 j_2}$，其中 W 的对角元素和相邻元素由 (14) 给出，其余非对角元素均为 a_k。若 $x_1 = C$ 或 $x_2 = C$，则令 $v(x,t) = a_k$。因此，$v(x,t) \le b$，且当且仅当 $x = t$ 时等号成立。这就证明了指定范围内每个 ε 的情形 (i)。

令 D^0 在收到报告 $(r_1,\ r_2) \in T$ 后，以概率 1 选择 $x = (r_1,\ r_2)$。在 M^+ 中，每个代理人 i 都可以发送 $T_i \cup \{C\}$ 中的一条消息，而消息 $(a_1,\ a_2)$ 会以概率 1 导致 $x = (a_1,\ a_2)$。当类型的边际分布为 p_ε 时，记 $S_\varepsilon := S(D^0)$。

断言 B.2. *F 中的每个成员都可在 D^0 的严格纯策略均衡中实现。特别地，$F \subseteq S_\varepsilon$，且 D^0 是真实的。*

证明。固定 $j \in \{0,\dots,k\}$，并让每个代理人的每种类型都报告 $(t_i + j) \bmod N$。固定一个代理人 i。另一个代理人以 $1-\varepsilon$ 的概率报告 j。令 $\mu_j := \mathbb{E}[x_{-i}]$ 为该代理人的期望报告。则 $|\mu_j - j| \le \varepsilon k < 1/2$。代理人 i 的每种类型 t_i 都可以通过报告 $(t_i + d) \bmod N$，选择任意差值 $d \in \{0,\dots,k\}$，从而获得期望效用 $-d^2 + 2d\mu_j - \mathbb{E}[x_{-i}^2]$。由于 $|\mu_j - j| < 1/2$，该期望效用在 $d = j$ 处取得唯一最大值。因此，指定的报告构成一个严格纯策略均衡。在每个类型组合下，两个差值都等于 j，因此委托人获得 $W_{jj} = a_j$。取 $j = 0$ 即可得到真实均衡。■

断言 B.3. *M^+ 具有唯一的贝叶斯纳什均衡，其中每种类型都如实地报告。*

证明。 进行迭代删除严格劣势策略：每一个虚假报告都被 C严格占优。排除它们之后，每种类型都严格偏好真实胜过 C。■

断言 B.4. $d_H(S_\varepsilon,\ F)\ \to 0$作为 $\varepsilon \downarrow 0$。

证明。 取任意序列 $\varepsilon_m \downarrow 0$ ，并且对每个 m，取相应先验分布下 D^0 的任意均衡 σ^m 。策略组合的空间是单纯形的固定有限乘积，因此存在一个子序列收敛到策略组合 σ^*。在任一代理人具有零类型的条件下，另一代理人的类型以趋近于一的概率为零。因此，将两个零类型代理人的最佳响应不等式取极限可知，$(\sigma_1^*(\cdot \mid 0)\ \sigma_2^*(\cdot \mid 0))$ 是完全信息博弈的一个均衡，在该博弈中，每个代理人从 $\{0,\ \ldots,\ k\}$ 中选择一个报告，并获得 $-(r_i - r_{-i})^2$。

证明 断言 A.1中的最近整数计算列举了该博弈的均衡。两个代理人要么选择相同的纯报告 j，要么都以相同的相邻报告 j，$j+1$等概率混合。委托人在第一种情况下获得 a_j，在第二种情况下获得 a_{j+1} ，依据（15）。由于真实策略组合$(0, 0)$的概率趋近于一，且委托人的收益独立于 ε而有界，委托人的期望收益收敛于 F中的同一个元素。

因此，最大值$_{s\in S_\varepsilon}$距离$(s,\ F)\ \to 0$。[21] 根据断言 B.2，F中的每个元素都 S_ε属于，因此最大值$_{a\in F}$距离$(a,\ S_\varepsilon)\ = 0$。■

现在，我选择一个 $\varepsilon > 0$指定范围内使得 $d_H\ (S_\varepsilon,\ F)\ < \eta/2$的，并固定由此得到的环境 E。结合断言 B.2，即可证明(ii)部分。

断言 B.5。 存在一个 $\delta > 0$ 使得（22）在此固定环境中成立。

证明。 反设不然。存在直接机制 D_m 满足

$$\max S(D_m) > b - \frac{1}{m}, \qquad \& \qquad d_H(S(D_m), F) \geq \eta.$$

[21]如果这一点不成立，就会存在一个均衡收益序列，与 F 保持正距离。选择实现这些收益的均衡并取一个收敛子序列，将与刚刚确立的极限相矛盾。

在每个机制中，选取一个 σ_m 达到最大值的均衡。直接规则空间和策略空间的紧致性保证，沿着某个子序列，$D_m \to D^*$ 和$\sigma_m \to \sigma^*$成立。将其代入（1）和（2）并取极限可知，σ^* 是 D^*的一个均衡，为委托人带来 b。断言 A.3 的证明中的论证同样适用于此处，因此D^* 是 D^0的重新标记。

$a \in F$ 的每个值都在 D^*的严格纯策略均衡处达到。有限个最佳响应不等式在直接规则充分接近D^*时仍然成立。因此，对于所有充分大的 m，同一个纯策略组合都是 D_m的均衡，其委托人收益收敛于 a。由于 F 是有限集，最大值$_{a\in F}$ dist$(a, S(D_m)) \to 0$。

由于 $d_H(S(D_m)\ F) \geq \eta$，对于所有充分大的 m，必然存在一个收益$s_m \in S(D_m)$，使得dist$(s_m, F) \geq \eta$。选择达到每个 s_m 的均衡，并从这些策略中进一步选取一个收敛的子序列。均衡不等式在极限处成立，因此收益收敛于某个 $s^* \in S(D^*) = S_\varepsilon$。但是，根据 ε的选择，dist$(s^*, F) < \eta/2$，这与dist$(s_m, F) \geq \eta$矛盾。由此证明(iv)部分。■

这四个结论在固定环境 E 中均成立。 ■

B.1 私人价值情形的刻画

定理 *3.1* 中关于私人价值断言的证明。 §4.1 中的证明表明，(iii) 蕴含 (ii)，适用于每个有限环境，包括具有私人价值和独立类型的环境。 (ii) 蕴含 (i) 在这一受限类别上也仍然成立。

为证明必要性，设对于某个紧集 S，V$(S) <$ 最大值 S。§4.1 中的有限逼近给出一个有限集 F，其中 V$(F) < b :=$ 最大值 F。令 $\gamma := (b - \text{V}(F))/2 > 0$，并选择 $\eta > 0$ 使得（9）成立。将引理 B.1 应用于这个 F 和 η，得到一个环境 E、一个机制 M^+ 和一个 $\delta > 0$。

如果某个直接机制满足 max $S(D) \leq b - \delta$，则由 (3) 可知 V$(S(D)) \leq b - \delta$。否则，由 (22) 和 (9) 可知 V$(S(D)) < b - \gamma$。因此，sup$^{D\in\mathcal{D}(E)}$V$(S(D)) \leq b - \min\{\delta, \gamma\} < b$。委托人的收益 $v(x, t)$ 至多为 b，而 M^+ 有一个使她获得 b的唯一均衡。因此，$V(E) =$ V$(S(M^+)) = b$，从而在

具有私人价值和独立类型的环境中。特别是，(i) 即使仅要求在这类环境中成立，仍不成立。■