第 7 卷第 1 期　　经 济 学（季 刊）　　Vol. 7, No. 1
2007 年 10 月　　China Economic Quarterly　　October, 2007

建立在行为经济学理论基础上的委托-代理模型：物质效用与动机公平的替代

蒲勇健*

摘　要　本文通过将 Rabin（1993）提出的同时考虑了物质效用和“动机公平”的效用函数植入现有的委托-代理模型，获得一个代理人表现出“互惠性”非理性行为的新委托-代理模型。由该模型给出来的最优委托-代理合约在一定条件下可以给委托人带来比现有委托-代理最优合约更高的利润水平。该研究改进和完善了蒲勇健（2007）的结果。与蒲勇健（2007）的研究相比，我们发现在同时考虑了物质效用和“动机公平”的情况下，非理性代理人的行为更加复杂，他会在公平与自利之间寻求平衡。数值模拟的结果表明，模型一般会存在解。

关键词　博弈论，信息经济学，行为经济学

一、现有委托-代理理论的不足

委托-代理关系广泛出现于各种经济管理问题中。大到国家对其下属各类机构的管理，中到行业管理机构与企业的关系，小到企业内部董事会与经理层，甚至工厂班组领导与其所管理的工人们，委托-代理关系几乎是无处不在，无时不有的。现有的委托-代理理论模型主要是来自信息经济学的直接应用，是博弈论在不对称信息情形的应用。该理论模型的主要思想是认为委托-代理关系中存在的主要问题是委托-代理合约签订之后可能会出现事后的信息不对称，即道德风险，从而损害对方的利益。代理人的道德风险会损害委托人的利益，而委托人的道德风险也会损害代理人的利益。现有的委托-代理模

* 重庆大学经济与工商管理学院，重庆大学发展研究中心，重庆大学“985 二期工程”哲学社会科学创新基地/欠发达地区可持续发展与循环经济研究中心。通讯地址：重庆市沙坪坝区沙正街 174 号重庆大学经济与工商管理学院，400044；电话：13983080099，(023)65472898；E-mail：puyjan@yahoo.com.cn。本文获得国家社会科学基金项目(07BJY017)的资助。本文得到多位匿名评论人的指导，为此特别向多位匿名评论人表示衷心感谢！还要感谢香港大学经济金融学院的赵耀华(Stephen Chiu)博士，是他首先将 Rabin1993 年的论文介绍给我才使得我开始接触行为经济学文献。他多次邀请我访问香港大学和香港中文大学，在我们十多年的学术合作中我从他那里索取甚多，受益匪浅。我对于我在我们俩的多年合作里分得较大的饼这种“低互惠”行为感到惭愧，并向他再次表示衷心的感谢！

型更多地对存在代理人道德风险情况下的激励机制进行了研究（Holmstrom，1979；Mirrlees，1975，1976；张维迎，1996；当然，对于存在委托人道德风险情况下的激励机制也有相关的研究，见张维迎，1996）。在委托人不能无成本地观察到代理人行为的情况下，委托人通过与代理人签订合约，约定按照某些双方都可观察到的指标体系来向代理人支付报酬（或进行奖惩）的办法对代理人激励。在这种模型中，委托人的支付函数是期望效用函数，委托人不能无成本地直接观察到代理人的行为或努力水平，而是通过激励相容约束条件对代理人的行为或努力水平进行间接的判断。委托人给予代理人的报酬一般分为两部分，一部分是固定报酬（如固定工资或基本工资），另一部分就是与业绩挂钩的变动报酬（或变动工资）。业绩一般用双方都可观察到的指标（体系）来衡量（如利润等）。在现在的主流模型中，最优激励机制或合约要求参与约束条件给予代理人的期望支付（或期望效用）恰好等于代理人的保留支付（或保留效用），或者代理人得到的期望支付比其保留支付多出点。这是委托人理性假定的必然结果。委托-代理模型近年来在已故的法国图卢兹学派领头人拉丰（Laffont）的推进下有新的发展（拉丰和马赫蒂摩，2002）。

现有的委托-代理理论模型存在这样的一个问题：代理人的行为只对其所获得的绝对收入水平作出反应，没有考虑代理人会关注其他人的收入水平；或者说只是假定代理人的效用仅仅取决于其个人绝对收入水平，没有考虑相对收入对于代理人效用的影响。[1]尽管在现有的委托-代理模型中也考虑了代理人相对业绩对于委托人设计最优合约的影响（即所谓的“标尺竞争”机制），但是代理人行为反应居然未考虑其相对收入不能不说一个研究上的严重缺失。[2]同时，代理人在现有的委托-代理模型里没有出现对于“公平”的追求也是与许多事实相左的。委托人在给予代理人的报酬分配中，如果代理人不止一个，就会出现代理人对于委托人进行的利益分配是否公平的关注，而这种关注一般会影响委托人的行为反应或努力程度的。相对收入效应一般又是与公平效应相联系的。我们可以从现实中观察到的现象来说明这一问题。

事实上，我们在现实中可以观察到许多与现有的委托-代理模型相矛盾的现象。譬如，在由 Fehr *et al*.（1993）所作的著名的“礼物交换博弈”（gift-exchange game）实验中，一些人扮演企业，他们给扮演工人的实验者发工

[1] 或许这里存在着“相对收入的心理影响是否能够量化，便于测量”的问题。事实上，微观经济学中的“效用”概念也存在着这样的测量问题。在这里，相对收入对于个人决策的影响，是来自个人的预期，而预期只需要引导个人的行为就行了，不要求一定要可以准确进行测量。

[2] 事实上，根据现有的文献，即使是多个委托人和代理人之间的关系，传统委托-代理模型都没有考虑相对收入效应。只有在最近一些年的行为经济学文献里才开始考虑相对收入效应。而本文是这种行为经济学研究领域中的一个贡献，也是沿着这样的思路对传统委托-代理模型进行批评的。所以，这里的批评不会是“用多个代理关系去批评一个代理关系”，而是来自行为经济学对于传统主流经济学假定的批评。

资。工人们接受工资后就选择自己的努力水平。选择的努力水平愈高，则企业的利润就愈大，而工人的效用就愈低。该博弈基本上是一个序贯“囚徒困境”博弈，其中工人们有一个占优策略是选择可能的最低努力程度。唯一的子博弈精练均衡工资就是保留工资。但是，实际的实验结果并非如子博弈精练均衡所预言的，而是显著高于保留工资，而工人的努力程度也显著高于其可能的最低努力程度。实验结果是工人的努力水平与工资水平呈正相关。所以，现有的委托-代理模型就不能解释这些现象。

笔者在研究中产生了一种想法，认为通过在现有的委托-代理理论模型中引入对于“公平”的考虑可以完善现有的委托-代理理论模型，并且还能够对上述研究结果加以解释。最近一些年来，许多研究者对于传统经济学关于理性人的假定提出异议，认为至少不是所有人的行为都能够用新古典经济学的效用最大化来加以说明的，而利他行为，即人们之间的相互帮助现象是普遍存在的。近年来出现的行为经济学也针对理性人假设进行修正，试图对于许多传统经济管理理论不能解释的现象予以解释。一项引人注目的原创性研究是 Rabin（1993）的思路，他在 J. Geanakoplos，D. Pearce 和 E. Stacchetti（文献中简称为 GPS）所提出的“心理博弈”框架基础上，构造了一个引入公平偏好的博弈论体系。他的关键性工作是通过对“公平”概念加以严密的定义来改造传统博弈论中的支付函数，从而发现了一些新的均衡。他得到的结果是，除了传统博弈论中已知道的纳什均衡之外，还出现新的“公平均衡”。这个工作还发现了许多“合作性均衡”，它们并不像传统博弈论那样要求无限次重复博弈或信息不对称条件。这种结果对利他行为和合作现象的解释是强有力的。Rabin（1993）的一个有价值的贡献是把“公平性”（fairness）定义为“当别人对你友善时你也对别人友善，当别人对你不善时你也对别人不善（即“投桃报李”和“以牙还牙”），而且他的独特贡献是将这种概念给予了明确的规定，即“如果你在损失自己效用（收入、利益等）的情况下去损害别人的效用（收入、利益等），就被定义为你对别人不善；如果你在损失自己效用（收入、利益等）的情况下去增进别人的效用（收入、利益等），就被定义为你对别人友善”。心理学的诸多实验证据表明，人的行为在许多情形下是遵循这样定义的“公平性”规则的，特别是在按照这种规则作出反应时所可能造成的潜在物质利益损失不太大的情况下更是如此。我们认为，按照 Rabin 的方法，可以解释为什么一些公司给予其职员的期望收入明显高于其保留收入。公司在给予其职员的收入明显高于他们的保留收入的情况下是损失公司自己的收入或利益，因而表现出对其职员的友善行为；而职员们基于“公平性”规则的考虑也在这种情况下对公司表现出友善；他们的友善行为可以表现在即使在没有公司管理人员的监督情况下也不会有偷懒或小偷小摸等损害公司利益的行为等。这样，公司给予其职员高于其保留收入的期望收入尽管减少了公司

的利润，但是由于职员们的“投桃报李”，公司可以通过减少监督而在监督费用上有节省。当节省的监督费用大于因为公司给予其职员高于其保留收入的期望收入而减少的公司利润时，公司总的利润水平却是增加的。所以，实际上这些公司的行为仍然是理性的，因为它们的行为仍然是追求利润最大化。运用 Rabin 的框架使得我们对于企业的多样化行为有了更深刻的认识。认识到企业可以将“公平性行为”作为新的要素加以运用来增加其利润。由于不同公司里的职员表现出不同程度的“公平性”行为，所以不同公司在将“公平性行为”作为新的要素加以运用来增加其利润时，给予其职员的固定收入一般是不同的，所以我们发现实际上每个行业里不同公司的人均固定收入都是不同的，并且高于保留收入。当然，根据现有的效率工资理论也可以对公司为何给予其职员高于他们的保留收入现象予以说明，但是这要求博弈是无限次重复的；而按照“公平博弈”的思路并不需要作出这样的假定。从这样的思路出发，我们可以将基于 Rabin 的公平性植入现有的委托–代理理论模型中，从而在维持现有的理性人假定下（即追求利润最大化假定）将现实中的公平偏好纳入最优机制或合约的设计里，完成对现有委托–代理理论模型的改造，并且获得同时考虑了理性人假定和公平性的一个新的委托–代理理论模型。这项研究的理论意义在于将行为经济学中有价值的要素与现有委托–代理理论模型相结合，从而完成对于现有委托–代理理论模型的修正和改造。这项工作的实用意义是其使得委托–代理理论的适用面更加广泛；按照这种理论模型所设计的最优激励合约也更加有效率，为企业带来更大的利润。这项研究还可能具有更加广泛的意义，因为我们所看到的不同企业在企业文化上的差异也许在某种意义上正是它们的管理层—职员在公平偏好上的差异造成的；而得到的预期理论模型还可能为我们在组织行为学和人力资源管理学（如雇佣决策和薪酬设计等）的研究中产生巨大的影响，甚至可能对现有的组织行为学和人力资源管理学带来新的创新性研究。

二、引入“公平博弈”的委托–代理模型：不存在物质效用效应的情形

本文的目的是在现有的委托–代理模型中引入“公平”要素，并且在对于“公平”概念的刻画上采用来自 Rabin（1993）的理论，即“互惠性动机公平”理论。Rabin 的“互惠”理论是属于行为经济学范畴的，因而本文工作的实质是在传统激励理论中引入行为经济学的假定。我们假定代理人具有 Rabin（1993）中所说的“互惠性”行为。对于委托人，我们在这里仍然假定其是传

统的理性人，这在委托人是公司或政府等非自然人的场合是恰当的。[3]这种假定可以使得我们去考虑这样的问题，即委托人如何利用代理人的非理性行为去设计工资制度，使得委托人获得最大利润，甚至比在现有委托–代理理论中（其中假定代理人是理性人）所获得的最大利润还要大的利润水平。

蒲勇健（2007）成功地将 Rabin 的思想植入了经典的委托–代理模型，并且发现其结果证明现有的 Holmstrom-Milgrom 模型中的最优合约不是帕累托最优的。

蒲勇健（2007）假定委托人知道代理人是非理性的，并且在行为上表现出这样一种“互惠性”——当委托人在牺牲自己利益的情况下给予代理人更多的利益时，他们会以情愿牺牲自己的利益来回报委托人的好意。他给出的具体刻画方法是，假定委托人给予代理人的固定收入要比现有委托–代理理论中恰好使得代理人获得等于其保留支付的固定收入要高一些。这样，委托人以牺牲一个等于这个高出保留支付的数量的利润给予代理人同样数量的收入增加。代理人的“投桃报李”行为假定表现为他们会在现有最优努力程度上再增加一定数量的额外努力，从而使得自己不能获得最大效用，但是可以增加委托人的利润，使得委托人的利润比在代理人选择传统委托–代理理论中的最优努力程度时所获得的利润还要多。代理人用这种方式以牺牲自己的效用为代价去增加委托人的利润。蒲勇健证明，在一定的条件下（这些条件对委托人和代理人各自牺牲自己利益的程度作出限制），代理人的非理性行为的确使得委托人得到的利润比代理人表现出理性行为时还要高，即使委托人这时付出了比现有委托–代理模型中的固定收入

[3] 对于非自然人的公司、企业和政府等，假定其行为为理性的是经济学中的一贯传统，而行为经济学和心理学实验也只是发现自然人有普遍的非理性行为，文献中并没有发现企业和政府也有非理性行为的记录，所以，这里的假定是十分自然的。也就是说，目前文献还没有记录实验中发现非自然人（企业、组织、政府等）有非理性行为，而通常在经济学研究中，委托人一般是非自然人（企业、组织、政府等）。论文中的假定是以实验结果为依据的，因此自然有这样的假定。这是该假定的第一个理由。第二个理由是，一般地，企业、组织、政府等的形成通常有特定的工作目标，如企业是以利润最大化为其目标，为实现其目标，其行为一般是理性的。另外，实际上，论文的结果并不受委托人是理性的假定影响。即使委托人也是非理性的，论文所获得的“互惠均衡”也是仍然成立的。论文之所以要假定委托人是理性的，除了上述考虑之外，主要是试图解释日本的企业文化。也就是说，我们认为理性的日本企业是有意利用其员工的非理性行为，施以恩惠从而获得员工的互惠反馈而获取更大利润。也就是说日本企业是通过这样的企业文化来追求更大利润。这反映在论文里假定委托人通过选择后面的参数 δ 来最大化其期望利润的数学处理。委托人对于其给予代理人的互惠程度 δ 是委托人可以选择调整的变量，就意味着委托人是为某种目的，论文当然假定委托人的目的就是追求期望利润的最大化。

当然，在其他的场合，委托人也可能是自然人。如一个人（委托人）委托其律师（代理人）为其打官司，两个人都是自然人。此时，可假定两个人都是非理性的。这时，委托人给予代理人的互惠程度 δ 就不是其任意调整的变量，而是根据不同的人有不同的固定数值（由其性格决定）。但即使在这样的场合，论文的结果也是基本成立的，只不过这时不能保证均衡合约一定是帕累托占优于 Holmstrom-Milgrom 模型给出的合约，但是，可以保证的是存在某些非理性委托人，使得均衡合约一定是帕累托占优于 Holmstrom-Milgrom 模型给出的合约；取决于委托人的非理性程度（即其互惠程度 δ）。只不过在委托人是理性人情况下，他可以对其互惠程度 δ 进行有意识的选择，使得自身的预期利润最大化而已。但论文的主要目的是研究企业或组织通过利用代理人的非理性行为获得更大利润这样的主题，所以假定两个人都是非理性的这样的情形不是论文所要考查的对象。

还要高的固定收入之代价。

下面先给出现有的 Holmstrom-Mirrlees 委托–代理理论的一种简化模型，由 Holmstrom and Milgrom（1987）（见张维迎，1996）给出，它是一个适当简化的一维连续变量一般化模型，用参数化方法表述。该模型在文献中被大量采用。设 a 为代表代理人程度的一维变量，生产函数为 $\pi=a+\theta$，π 是产出或利润，θ 是均值为零、方差为 σ^2 的正态分布随机变量，是外生不确定因素。故有

$$\mathrm{E}\pi = a, \quad \mathrm{Var}(\pi) = \mathrm{Var}(a) + \mathrm{Var}(\theta) = \sigma^2,$$

其中 E 表示数学期望算子，Var 表示方差。

假定　委托人是风险中性的，代理人是风险规避的。

现在考虑线性合约（当委托人和代理人的绝对风险规避度都为常数时，理性代理人情形下的最优合约是线性的，这里假定不变的绝对风险规避度；（见张维迎，1996）：

$s(\pi)=\alpha+\beta\pi$，α 为代理人的固定收入，β 为代理人分享的产出份额，$s(\pi)$是代理人的收入。

因为委托人是风险中性的，则给定 $s(\pi)=\alpha+\beta\pi$，设委托人的效用函数是 $v(\pi-s(\pi))$，

代理人的绝对风险规避度为 ρ，代理人的努力成本 $c(a)$等价于货币成本，Holmstrom 与 Milgrom 具体设为

$c(a)=\dfrac{ba^2}{2}$，这里 $b>0$ 为成本系数；

设 $\bar{\omega}$ 为代理人的保留收入水平，则当代理人的确定性等价收入 $x<\bar{\omega}$，理性代理人将不接受合约。故参与约束可用确定性收入表示为

$$x = \alpha + \beta a - \frac{1}{2}\rho\beta^2\sigma^2 - \frac{b}{2}a^2 \geqslant \bar{\omega}.$$

现在看 a 是委托人不可观测时（信息不对称时）的最优合约。

理性委托人的最优合约是满足下面问题的解（最大化其确定性收入）（见张维迎，1996））

$$\max_{\alpha,\beta}[-\alpha+(1-\beta)a],$$

$$\text{s. t. (IR)}\quad \alpha + \beta a - \frac{1}{2}\rho\beta^2\sigma^2 - \frac{b}{2}a^2 \geqslant \bar{\omega},$$

$$\text{(IC)}\quad a = \frac{\beta}{b}.$$

在 Holmstrom-Milgrom 的模型里，个人理性约束（IR，其中不等式的左端是代理人的确定性等价收入）为等式，这是理性委托人假定的必然推论。

因为可以假定理性代理人在与委托人签订合约和不签订合约之间是无差异的情况下，代理人总是选择签订合约。此时，委托人的最优固定工资设计为使得个人理性约束为等式。于是有 $\alpha=\bar{\omega}-\beta a+\frac{1}{2}\rho\beta^2\sigma^2+\frac{b}{2}a^2$，将 IR 和激励相容约束 IC 条件代入目标函数，得：

$$\max_{\beta}\left[\frac{\beta}{b}-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}\left(\frac{\beta}{b}\right)^2-\bar{\omega}\right],$$

一阶条件为：$\frac{1}{b}-\rho\beta\sigma^2-\frac{\beta}{b}=0$，

得 $\beta=\frac{1}{1+b\rho\sigma^2}>0$，

$\alpha=\bar{\omega}+\frac{1}{2}\left(\rho\sigma^2-\frac{1}{b}\right)\frac{1}{(1+b\rho\sigma^2)^2}$，这就是当委托人和代理人都是理性人时的最优合约。

此时委托人得到的期望收入为

$$\mathrm{E}v=-\bar{\omega}+\frac{1}{2b(1+b\rho\sigma^2)}.\tag{1}$$

蒲勇健（2007）假定代理人是非理性的，其行为遵循前面所刻画的“互惠性”。设委托人给予代理人的固定工资高于前面在 Holmstrom-Milgrom 模型里规定的水平，使得代理人在最优努力水平下的确定性收入高于其保留支付。设这个固定工资差为 δ。假定代理人的“互惠性”反应是选择比在理性假定下的最优努力程度更加努力地工作，这里假设为代理人多付出工作努力量 a^*。不同的文化环境、宗教信仰会引至不同的互惠性行为。所以在这里不同文化和宗教背景下的代理人所多付出工作努力量 a^* 也是因人而异的。此时代理人的确定性收入为 $\bar{\omega}+\gamma$。一般可假定有 $0\leqslant\gamma\leqslant\delta$，这是“互惠性”的要求。设在理性代理人情形里委托人设计的最优固定工资为 α_0，于是有

$$\alpha_0+\delta+\beta(a+a^*)-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}(a+a^*)^2=\bar{\omega}+\gamma,\tag{2}$$

解得

$$a^*=\frac{1}{b}\sqrt{2b(\delta-\gamma)+(1-b\rho\sigma^2)\left[\beta^2-\frac{1}{(1+b\rho\sigma^2)^2}\right]}\text{（舍去负根）}.$$

蒲勇健（2007）在这里只考虑理性委托人仅改变固定收入的情形（本文也是如此），在新的合约里的绩效工资因子与原有委托-代理合约中相同，于是有 $\beta=\frac{1}{(1+b\rho\sigma^2)}$，所以有

$$a^{*}=\frac{1}{b}\sqrt{2b(\delta-\gamma)}.$$

在前面的假定下，有 $\gamma\leqslant\delta$，所以这个解是有意义的实数。当 $\gamma<\delta$ 时，$a^{*}>0$，意味着总产出较之代理人理性行为下的最优产出有所增加。但在 $0\leqslant\gamma<\delta$ 时，由式（2）知道，此时代理人的收入与其理性的努力选择下的收入相比较下降了，因而此时委托人的收入必然是增加了的。所以，代理人的非理性行为一定是“互惠性”的。

在非理性代理人情形，委托人的期望收入为

$$\begin{aligned}\mathrm{E}v^{*}&=-(\alpha_0+\delta)+(1-\beta)(a+a^{*})\\&=-(\alpha_0+\delta)+(1-\beta)\left(\frac{\beta}{b}+\frac{1}{b}\sqrt{2b(\delta-\gamma)}\right).\end{aligned}\tag{3}$$

理性委托人选择 δ 最大化其期望收入，于是有下列一阶条件：

$$\frac{\partial\mathrm{E}v^{*}}{\partial\delta}=-1+\frac{(1-\beta)}{\sqrt{2b(\delta-\gamma)}}=0,$$

得到

$$\delta=\gamma+\frac{(1-\beta)^2}{2b}=\gamma+\frac{b\rho^2\sigma^4}{2(1+b\rho\sigma^2)^2},$$

显然有 $\delta\geqslant\gamma$。

委托人的最大期望收入为

$$\begin{aligned}\mathrm{E}v^{*}&=-\bar{\omega}-\gamma+\frac{1}{2b(1+b\rho\sigma^2)}+\frac{b\rho^2\sigma^4}{2(1+b\rho\sigma^2)^2}\\&=\mathrm{E}v-\gamma+\frac{b\rho^2\sigma^4}{2(1+b\rho\sigma^2)^2},\end{aligned}\tag{4}$$

这里用到了式（1）。

当 $\frac{b\rho^2\sigma^4}{2(1+b\rho\sigma^2)^2}\geqslant\gamma$，有 $\mathrm{E}v^{*}\geqslant\mathrm{E}v$，此时委托人利用代理人的非理性行为就可获得比按照现有委托-代理最优合约所获得的期望收入还要高的期望收入，这说明，如果有 $\rho\neq0$，$\sigma\neq0$（一般场合是如此的），只要代理人的互惠性动机足够强烈（即 γ 足够小），此时就有 $\frac{b\rho^2\sigma^4}{2(1+b\rho\sigma^2)^2}\geqslant\gamma$，委托人利用代理人的非理性行为经济原理设计其工资合约就是有价值的。[4] 蒲勇健（2007）指出，这说明在一定条件下，企业对其职员更加人性化的关怀，给予员工比其保留支付还要多的固定收入不仅不会减少企业的利润（本文中支付，效用、利润

[4] 注意，在代理人是非理性的假定下，这种机制是可自我实施的，这种互惠性是稳定的。

和收入等概念是在可互换意义上使用的），而且正好相反，企业因此而激发起企业员工的感激之情，使得员工更加努力地为企业工作。[5]

这个模型可以解释许多成功企业的人性化管理和相应的企业文化，可以解释如日本企业里员工在下班后还仍然在为企业无偿工作等现象，也可以解释大公司给予其高管高年薪收入的现象。在日本的许多企业里，公司给予员工非常好的福利，公司高层也经常在业余时间里与员工们进行平等友好的交流沟通；反过来，员工的回报是在公司处于困境时也不像欧美企业中的员工那样频频跳槽，而是宁愿接受减薪而在原来的公司里继续努力工作，与公司共渡难关。这种现象在韩国企业中也能够看到。这都是通过这个模型可以得到解释的现象。尽管在日本和其他的一些亚洲企业里，早就有互惠性的激励制度了。但是，我们也观察到，这种互惠性的激励制度在包括欧洲在内的其他地方并不显著。欧美企业里的员工普遍表现出给钱才出力，绝不主动为企业免费打工的现象，同时企业也给员工开出按照边际产值计算的薪水。蒲勇健（2007）认为这个问题有几种可能的解释。首先，按照行为经济学中著名的最后通牒博弈实验、独裁博弈实验、礼物交换博弈实验和公共品博弈实验，互惠行为只是多数，并不是所有人都表现出互惠行为，它是统计上显著的。因此，理性人实际上也是存在的。其次，欧美企业里强大的工会组织是劳动市场里的垄断者。正如在前面已经假定和说明的那样，作为一个组织通常是理性的，尽管可以认为其中的大多数个人是非理性人，其所在组织一般是理性人。所以，工会与资方谈判决定的工资合约就是理性人之间形成的合约，没有互惠性；最后，Rabin 的“互惠”理论中有多重均衡，“互损”也是均衡。有人就用这种负的“互惠”即“互损”均衡来解释非志愿失业（见 Akerlof，1982；Akerlof and Yellen，1988）。这样的解释就意味着欧美企业中员工与老板之间通常出现的是“互损”均衡，而包括日本和韩国在内的一些亚太地区企业文化中通常出现的是员工与老板之间的“互惠”均衡。当然，这只是一种可能的解释。因为这种解释意味着作为理性人的欧美企业资方不知道大多数员工是非理性的，可以通过设计互惠性工资合约来获得帕累托改善。蒲勇健（2007）认为，第二种解释是最恰当的，因为欧美企业中的工会组织远比东亚地区企业里的工会组织强大和完善是一个不争的事实。

---

[5] 本文中已经证明：在互惠均衡下委托人和代理人都较理性人假定下的委托-代理合同下获得的效用高。因为有 $Ev^{*}\geqslant Ev$ 和 $0\leqslant\gamma\leqslant\delta$。注意，这里所说的效用是净效用。在模型里，代理人在委托人并没有增加监督（从而没有增加其监督成本）的情况下，为回报委托人的善意而单方面增加其努力程度就已经包含有代理成本的节约的意思了。所以，代理人产出增量和代理成本的节约一定会大于多支付给代理人的工资，这样的结论是隐含在模型分析之中的。另外，在委托-代理模型里，已经不把委托人的监督作为变量加以考虑了。之所以委托人要与代理人之间签订委托-代理合约，是因为在边际上委托人的边际监督收益已经小于其边际监督成本。所以，在讨论委托-代理合同时，已经有这样的潜在假定。因此，本文不同时研究监督问题和委托-代理合同问题。在本文里，谈到互惠性会节约委托人的监督成本，是从一般情况来说的。在这里的模型里，我们并没有把监督成本变量显性地表达出来。它只是隐含于模型里（隐含的方式在上面已经谈到）。

根据式（2），当 $\delta>0$，$\gamma>0$ 时，基于代理人的这种非理性“互惠性”行为的工资合约给委托人和代理人都带来比 Holmstrom-Milgrom 合约更高的支付，所以，新合约带来的是一种帕累托改进。这种帕累托改进在理性代理人的 Holmstrom-Milgrom 框架里是无法实现的。这也说明，Holmstrom-Milgrom 模型中的最优合约不是帕累托最优的。蒲勇健（2007）猜想，包括文化在内的一些潜制度，如习俗、宗教等可能都有将这些仅靠理性合约不能够实现的帕累托最优加以实现的功能，这种猜想有待于进一步的研究加以验证或者证伪。

可以将 δ 理解为委托人对代理人的“善意程度”，代理人在委托人给予其较大的“善意程度”的情况下，理应也回报以较大的“善意程度”，这是人之常情。所以，蒲勇健（2007）自然假定 γ 是 δ 的函数。当 δ 愈大时，γ 就应该愈小；所以有 $\gamma=\gamma(\delta)$，$\gamma'(\delta)<0$。

委托人在式(3)中选择 δ 使得 Ev^* 最大化。一阶条件为：

$$\frac{\mathrm{d}Ev^*}{\mathrm{d}\delta}=-1+\frac{(1-\beta)(1-\gamma')}{\sqrt{2b(\delta-\gamma)}}=0,$$

解得

$$\delta=\gamma(\delta)+\frac{b\rho^2\sigma^4(1-\gamma')^2}{2(1+b\rho\sigma^2)^2},$$

代入式(3)，由式（1）得到

$$\mathrm{E}v^*=\mathrm{E}v-\gamma+\frac{b\rho^2\sigma^4[2|1-\gamma'|-(1-\gamma')^2]}{2(1+b\rho\sigma^2)^2},$$

如果 γ 是常数，则 $\gamma'=0$，上式就回到式（4）。

当且仅当 $\gamma\leqslant\frac{b\rho^2\sigma^4[2|1-\gamma'|-(1-\gamma')^2]}{2(1+b\rho\sigma^2)^2}$ 时，有 $\mathrm{E}v^*\geqslant\mathrm{E}v$；

当且仅当 $|\gamma'|<1, 0<\frac{b\rho^2\sigma^4[2|1-\gamma'|-(1-\gamma')^2]}{2(1+b\rho\sigma^2)^2}$；此时存在 γ 的可能性满足所有的条件。事实上，函数 $\gamma=-c\delta+d$ 就满足这些条件，其中 c，d 为常数，满足 $0<c<1$，$\frac{d}{c}\geqslant\delta\geqslant\frac{d}{c}-\frac{b\rho^2\sigma^4[2|1-c|-(1-c)^2]}{2c(1+b\rho\sigma^2)^2}=\frac{d}{c}-\frac{b\rho^2\sigma^4[1-c^2]}{2c(1+b\rho\sigma^2)^2}$。

这样可以保证有

$$0\leqslant\gamma\leqslant\frac{b\rho^2\sigma^4[2|1-\gamma'|-(1-\gamma')^2]}{2(1+b\rho\sigma^2)^2}.$$

其实，为了刻画代理人的“互惠性”非理性行为，不一定要求 γ 是 δ 的

函数。根据式（3），即使 γ 是常数（此时 γ 是 δ 的常数函数），因为 δ 愈大时 a^* 也愈大，代理人牺牲的自身利益也愈多，所以代理人的“互惠性”非理性行为也完全可以通过 a^* 来刻画。这可以通过 a^* 的表达式 $a^*=\frac{1}{b}\sqrt{2b(\delta-\gamma)}$ 看出。

一个特别的情形是代理人对于委托人的“善意”给予完全“对等”的回应，此时代理人把委托人给予他的超出 Holmstrom-Milgrom 合约所规定的支付更高的支付完全“还”给委托人，即有 $\gamma=0$，即代理人总是使得自己只得到保留支付。一般地，如果 δ 足够大，γ 当然是会等于 0 的。此时参与约束是紧的。这是理论上的特例，不是一般情况。甚至，在特别情况下，如果 δ 足够大，γ 还可能为负！在苏联初期，由于老百姓感激共产党给他们带来了平等和解放，老百姓在本来应该是休息的星期六下午自愿去工厂劳动，即所谓的“星期六义务劳动”，是没有报酬的自愿性劳动，此时 γ 就是负的。

蒲勇健认为在理论上存在着许多可想象出来的刻画代理人的“互惠性”非理性行为的方式，并且指出这些不同方式所形成的合约性质是下一步工作要研究的问题。当然，不同的人所呈现出来的互惠程度可能是不同的，并且对于委托人来说代理人的互惠程度是其所不知道的“私人信息”。委托人在设计合约时面临关于代理人互惠程度的不确定问题。委托人可以通过“试错法”逐步了解代理人的真实互惠程度来完善其合约。这种可能也适用于这样一种情况，即当代理人是互惠程度最低的即理性人时［注意，在蒲勇健（2007）的模型中从非理性人到理性人的变化在数学刻画上是连续的］，委托人就设计出通常在欧美企业里所看到的非互惠合同。所以，即使代理人的“道德修养”很低，委托人按照这样的原则设计合约也不会导致亏损。在现代企业中，老板因人而异地设计劳动合同是司空见惯的。

蒲勇健（2007）通过引入行为经济学方面的理念，并且在将这种理念加以具体刻画的基础上，改造了经典委托-代理理论中的 Holmstrom-Milgrom 模型，使之能够在代理人的行为动机上包含“互惠性”非理性行为在内的比单纯的理性人假定更加丰富的内容，而在现有的理性人假定下的委托-代理理论即 Holmstrom-Milgrom 模型只是蒲勇健所发现的模型之特例。譬如，在蒲勇健得到的模型里令代理人是理性的，则给定委托人的任何 δ，代理人选择的努力程度都是最大化式（2）的左端；容易验证，此时由蒲勇健的模型给出来的最优合约就正好是 Holmstrom-Milgrom 模型。

由于 Holmstrom-Milgrom 模型是现有激励理论和整个信息经济学的最基本的模型，激励理论和信息经济学是从这个模型框架发展而来的，所以蒲勇健（2007）所作的工作具有开创性的意义。因为，可以通过这种经过行为经济学理念的植入，去完成对整个信息经济学和激励理论体系的改造。这种改造是需要的，因为蒲勇健已经发现，至少对于经典的 Holmstrom-Milgrom 模

型来说，其最优激励合约是非帕累托最优的。也就是说，从蒲勇健所得到的模型出发，可以开展对于现有信息经济学和激励理论基本框架的重建。这将是未来的一个系统且漫长但又是激动人心的系列工作。

蒲勇健（2007）指出，沿着这样的思路可以对文化习俗甚至宗教行为等“潜制度”加以“经济学”解释。文化和宗教活动要耗用资源。文化和宗教活动并不一定是人类的需要已经获得相当满足的情况下才出现的。事实上，在人类的生产力水平和需要满足水平十分低下的远古甚至在原始社会，都有相当规模的文化和宗教活动。我们现在在许多考古现场和远古遗迹中不难证实这一点。如果理性人之间常常不能够取得帕累托最优的合作效果，文化和宗教中的圣贤就会教诲人们要对爱你的人表现出“投桃报李”，不能“忘恩负义”，“吃水不忘挖井人”，而对于有害于你的人要“以眼还眼，以牙还牙”，“一报还一报”等诸如此类的“互惠”型（“互损”型）行为（其他的还有“己所不欲，勿施于人”）。这样就使得人们在这种“互惠”型行为里有可能取得帕累托最优的“双赢”合作效果。从这种意义上说，文化和宗教可能具有提升社会生产效率的效果，因而也是“经济活动”。与我们熟知的经济活动一样，文化和宗教活动要耗用资源（包括自然资源和人力资源），也同时提供“经济产出”——带来社会交往（包括经济交易）中的帕累托改进。当然，不同的文化活动和宗教说教在这方面所能够取得的效果是不同的，但这也说明了这样的可能性是存在的。这种理论有可能为文化和宗教的起源和功能提供一种来自行为经济学的解释。另外，这样的模型有可能为“和谐社会”的建设提供政策上的微观经济理论基础，因为正是劳资双方和上下级之间或者更加一般地是社会成员之间的“互惠”，使得合作得以实现，既提高了经济效率又实现了“和谐”，从某种意义上看，“互惠”就是“和谐”。关于这些问题的后续研究将会是一个广阔的学术天地，这样的思路为沟通经济学与包括文化和宗教学在内的一般意义上的社会科学研究在方法论上架构了一座桥梁。

三、同时考虑物质效用效应和动机公平的委托-代理模型

蒲勇健（2007）的工作有一个局限性，即没有考虑代理人的物质消费效用，或者说没有考虑代理人在追求公平的同时所在物质效用损失上所付出的代价。事实上，即使考虑公平性，代理人的效用函数中也要考虑自利性的物质效用。作为比较全面的分析，模型中应该考虑代理人在自利性物质效用与公平感之间的权衡。Rabin（1993）那篇文章里，就给出了一种同时包括物质效用与公平追求的效用函数。我们在本文里，将进一步完善蒲勇健（2007）的工作，通过引入 Rabin 动机公平效用函数，获得更加全面的结果。

下面我们首先给出 Rabin（1993）动机公平模型。

考虑一个有局中人 1 和 2 的战略式二人博弈，它具有从有限纯战略空间 A_1 和 A_2 中导出的（混合）战略空间 S_1 和 S_2。设 π_i：$S_1 \times S_2 \to R$ 是局中人 i 的物质性收入（material payoffs）。

根据这个“物质性博弈”，Rabin 构造了一个“心理博弈”（psychological game），当局中人 i 选择其战略是依赖于下列三个因素时，它由预期效用来定义：（1）他的战略；（2）他关于其他局中人战略选择的信念；（3）他关于其他局中人关于他自己战略选择的信念的信念。

我们将使用下列符号：$a_1 \in S_1$ 和 $a_2 \in S_2$ 表示两个局中人所选择的战略；$b_1 \in S_1$ 和 $b_2 \in S_2$ 分别表示局中人 2 关于局中人 1 所选择战略的信念，以及局中人 1 关于局中人 2 所选择战略的信念；$c_1 \in S_1$ 和 $c_2 \in S_2$ 表示局中人 1 认为局中人 2 认为局中人 1 所选择战略的信念的信念，以及局中人 2 认为局中人 1 认为局中人 2 所选择战略的信念的信念。

Rabin 将公平性植入分析的第一步是要定义一个“友善函数”（kindness function）$f_i(a_i, b_j)$，它测度局中人 i 对于局中人 j 的友善程度。这个模型要假定两个局中人都具有相同的友善及公平的概念，并且对称地运用这些标准。Rabin（1993）曾经证明，如果允许多种不同的友善函数，该论文中得到的大多数结果都仍然是成立的。

如果局中人 i 认为局中人 j 选择了战略 b_j，局中人 i 选择 a_i 会有多么友善呢？倘若局中人 j 选择了战略 b_j，局中人 i 就从所有可行的收入集合中［即从集合 $\Pi(b_j) \equiv \{(\pi_i(a, b_j),\ \pi_j(b_j, a) \mid a \in S_i)\}$ 中］选择了这样一个收入组合 $(\pi_i(a, b_j), \pi_j(b_j, a)) \mid a \in S_i)$。局中人们会对于局中人 i 在 $\Pi(b_j)$ 中选择任何给定点产生许多不同的有关局中人 i 是多么友善的概念。Rabin 用一种特定的（是有意简单一些的）友善测度来进行下面的研究。

设 $\pi_j^h(b_j)$ 为局中人 j 在 $\Pi(b_j)$ 中的最高收入，并且设 $\pi_j^\ell(b_j)$ 是局中人 j 在 $\Pi(b_j)$ 中的帕累托前沿上的最低收入。设“等收入”（equitable payoff）为 $\pi_j^e(b_j) = [\pi_j^h(b_j) + \pi_j^\ell(b_j)]/2$。当帕累托前沿是线性的时，如果局中人 i 与其在帕累托前沿点上“平分好处”（split the difference），这个收入就对应于局中人 j 会得到的收入。更为一般地，它提供了一个初步的参照点，根据它可以测度局中人 i 对待局中人 j 是多么样的友善。最后，设 $\pi_j^{\min}(b_j)$ 是局中人 j 在集合 $\Pi(b_j)$ 中最糟糕的可能收入。

根据这些收入，Rabin 定义了友善函数。该函数刻画了局中人 i 认为他给予局中人 j 的有多少是大于或小于局中人 j 的等收入。

定义 1　局中人 i 对于局中人 j 的友善为

$$f_i(a_i, b_j) \equiv \frac{\pi_j(b_j, a_i) - \pi_j^e(b_j)}{\pi_j^h(b_j) - \pi_j^{\min}(b_j)},$$

若 $\pi^h(b_j) - \pi_j^{\min}(b_j) = 0$，则 $f_i(a_i, b_j) = 0$。

注意，$f_i=0$ 当且仅当局中人 i 打算将等收入给予局中人 j。如果 $f_i<0$，则局中人 i 给予局中人 j 的要少于其等收入。回忆一下等收入的定义，一般存在着两种可能使得 f_i 为负数，或者局中人 i 攫取的多于其在 $\Pi(b_j)$ 中的帕累托前沿上的份额，或者他选择了 $\Pi(b_j)$ 中的无效率点。最后，如果局中人 i 给予局中人 j 的要多于其等收入，则 $f_i>0$。回忆一下，只有在 $\Pi(b_j)$ 的帕累托前沿是非单点集时才会如此；否则，有 $\pi_j^e=\pi_j^h$。

Rabin 用函数 $\tilde{f}_j(b_j,c_i)$ 来表示局中人 i 关于局中人 j 是怎样友善地对待他自己的信念。在继续保持这两种符号不同的情况下，这个函数在形式上等价于函数 $f_j(a_j,b_i)$。

定义 2　局中人 i 关于局中人 j 是怎样友善地对待他自己的信念定义为

$$\tilde{f}_j(b_j,c_i)\equiv\frac{\pi_i(c_i,b_j)-\pi_i^e(c_i)}{\pi_i^h(c_i)-\pi_i^{\min}(c_i)},$$

若 $\pi_i^h(c_i)-\pi_i^{\min}(c_i)=0$，则 $\tilde{f}_j(b_j,c_i)=0$。

因为友善函数是正规化的（normalized），$f_i(\cdot)$ 和 $\tilde{f}_j(\cdot)$ 的值就必定位于区间［-1，1/2］中。进一步地，友善函数对于物质性收入的正仿射变换是不敏感的（对于即将定义的总效用来说，它对这样的变换将是敏感的）。

这样的友善函数就能够用于充分地刻画局中人偏好。每个局中人 i 选择 a_i 最大化其期望效用 $U_i(a_i,b_j,c_i)$，它融合了其物质性效用及局中人们共同认可的公平概念：

$$U_i(a_i,b_j,c_i)\equiv\pi_i(a_i,b_j)+\tilde{f}_j(b_j,c_i)\cdot[1+f_i(a_i,b_i)].$$

这样的偏好之主要行为特征反映出了我们最初的讨论内容。

如果局中人 i 认为局中人 j 对他不善——$\tilde{f}_j(\cdot)<0$——则局中人 i 也打算对局中人 j 不好，方法是选择一个行动 a_i，使得 $f_i(\cdot)$ 较低或为负。倘若局中人 j 善待局中人 i，则 $\tilde{f}_j(\cdot)$ 为负，且局中人 i 打算好好地对待局中人 j。当然，这种特定的效用函数是局中人在他们对于公平性的偏好与其物质利益之间进行权衡（trade off），并且对于物质的追求可能胜于对公平性的关注。

由于友善函数存在着上下限，这个效用函数体现了典型事实 C：物质性收入愈大，则局中人的行为中反映出他们对于公平性的关注就愈小。因此，在这些博弈中的行为对于物质性收入的大小规模是敏感的。

由于这些偏好形成一个心理博弈，就可以利用由 GPS 所定义的心理纳什均衡（psychological Nash equilibrium）概念；这只是纳什均衡概念在心理博弈中的一种类比，它增加了一些额外的条件即所有的高阶信念（higher-order beliefs）都与实际的行为相合。Rabin 将称如此定义的解概念为“公平均衡”。

定义 3　战略组合$(a_1, a_2) \in (s_1, s_2)$是一个公平均衡，如果对于 $i=1,2, j\neq i$，有

（1）$a_i \in \arg\max_{a\in S_1} U_i(a, b_j, c_i)$；

（2）$c_i = b_i = a_i$。

运用 Rabin 的这个动机公平模型，我们下面就来构造同时考虑物质效用和动机公平的委托-代理模型。

假设 $A>0, \overline{\alpha}>0, b>0, \rho>0, \sigma>0, \overline{\omega}>0, \beta=\dfrac{1}{1+b\rho\sigma^2}, 0\leqslant a\leqslant A, 0\leqslant \alpha\leqslant \overline{\alpha}$. 其中 A 是代理人可能付出的最大努力程度（由代理人的生理极限决定）；α 是委托人付给代理人的固定工资，而 $\overline{\alpha}$ 是委托人付给代理人的可能的最大固定工资（由委托人的融资能力决定）。按照 Rabin 的上述理论及定义 1 和 2，我们有（这里的 i，j 分别表示代理人和委托人）

$$\pi_i(a,\alpha) = -\exp\left[-\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2 a^2-\frac{b}{2}a^2\right)\right],$$

$$\pi_j^h(\alpha) = -\alpha+(1-\beta)\,a\,\big|_{a=A} = -\alpha+(1-\beta)A,$$

$$\pi_j^l(\alpha) = -\alpha,$$

$$\pi_j^{\min}(\alpha) = -\alpha,$$

$$\pi_j(\alpha, a) = -\alpha+(1-\beta)a,$$

$$\pi_j^e(\alpha) = -\alpha+\frac{1-\beta}{2}A,$$

根据定义 1 有

$$f_i(a,\alpha) = \frac{a}{A}-\frac{1}{2},$$

我们继续有

$$\pi_i^h(a) = -\exp\left[-\rho\left(\overline{\alpha}+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right],$$

$$\pi_i^l(a) = -\exp\left[-\rho\left(\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right],$$

$$\pi_i^e(a) = -\frac{1}{2}\exp\left[-\rho\left(\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]\left[1+\exp(-\rho\overline{\alpha})\right].$$

我们只考虑公平均衡的情形，此时根据定义 3，定义 2 中的 $c_i, b_i, a_i(=a)$ 满足均衡条件 $c_i=b_i=a_i$。根据定义 2 有

$$\tilde{f}_j(\alpha, a) = -\frac{1}{2}+\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\overline{\alpha})},$$

于是有

$$U_i(a,\alpha)=-\exp\left[-\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]$$
$$+\left[-\frac{1}{2}+\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}\right]\left[1+\frac{a}{A}-\frac{1}{2}\right]$$
$$=-\exp\left[-\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]$$
$$\cdot\left[\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}-\frac{1}{2}\right]\left[\frac{1}{2}+\frac{a}{A}\right],$$

代理人选择 a 以最大化自己的效用函数 $U_i(a,\alpha)$，一阶条件是

$$\frac{\partial U_i}{\partial a}=\rho\exp\left[-\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right](\beta-ba)$$
$$+\frac{1}{A}\left[\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}-\frac{1}{2}\right]=0,$$

或者

$$a=\frac{\beta}{b}-\frac{1}{A\rho b}\left[\frac{1}{2}-\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}\right]\exp\left[\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right],\tag{5}$$

当且仅当 $\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}<0$ 即 $e^{-\rho\alpha}<\frac{1+e^{-\rho\bar{\alpha}}}{2}$ 时有 $a>\frac{\beta}{b}$，也即当且仅当 $0<\alpha^*=-\frac{1}{\rho}\log\left(\frac{1+e^{-\rho\bar{\alpha}}}{2}\right)<\alpha\leqslant\bar{\alpha}$ 时（不难验证对于任意的 $\bar{\alpha}>0$ 有 $\alpha^*=-\frac{1}{\rho}\log\left(\frac{1+e^{-\rho\bar{\alpha}}}{2}\right)<\bar{\alpha}$）有 $a>\frac{\beta}{b}$。

所以当 $\alpha^*<\alpha\leqslant\bar{\alpha}$ 时有 $a>\frac{\beta}{b}$。由于 $a=\frac{\beta}{b}$ 正好是理性代理人（以及理性委托人）情形的最优委托–代理合同下的均衡能力程度，所以在非理性代理人情形下当 $\alpha^*<\alpha\leqslant\bar{\alpha}$ 时，代理人的努力程度要比理性代理人情形下要高。

由于

$$\frac{\partial a}{\partial \alpha}=\frac{\dfrac{\exp\left[\rho\left(\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]}{\left[1-\exp(-\rho\bar{\alpha})\right]Ab}-\dfrac{\left[\frac{1}{2}-\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}\right]}{A\rho b}\exp\left[\rho\left(\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]}{1+\dfrac{\left[\frac{1}{2}-\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}\right]}{A\rho b}\exp\left[\rho\left(\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right](\beta-ba)},$$

当 $\alpha>\alpha^*$ 时有 $\frac{\partial a}{\partial\alpha}>0$，此时 a 是 α 的增函数，即委托越是慷慨地给予代理人更多的固定收入，代理就越努力。

将式（5）代入委托人的期望效用函数 $U_j=-\alpha+(1-\beta)a$，得到

$$U_j=-\alpha+\frac{(1-\beta)\beta}{b}-\frac{(1-\beta)}{A\rho b}\left[\frac{1}{2}-\frac{1-\exp(-\rho\alpha)}{1-\exp(-\rho\bar{\alpha})}\right]$$

$$\cdot \exp\left[\rho\left(\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right)\right]. \tag{6}$$

我们关心在什么时候非理性代理人合同要给委托带来比理性代理人合同更加高的期望利润，即要考察在什么情况下有 $U_j > -\bar{\omega}+\dfrac{1}{2b(1+b\rho\sigma^2)}$，即

$$\bar{\omega}-\alpha+\frac{(b\rho\sigma^2-1)}{b(1+b\rho\sigma^2)^2}-\frac{\sigma^2}{(1+b\rho\sigma^2)A}\left[\frac{1}{2}-\frac{1-\mathrm{e}^{-\rho\alpha}}{1-\mathrm{e}^{-\rho\bar{\alpha}}}\right]$$
$$\cdot \exp\left\{\rho\left[\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right]\right\}>0, \tag{7}$$

同时代理人的参与约束为

$$\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2 \geqslant \bar{\omega}. \tag{8}$$

这里的数学问题是同时满足式（7），（8）的解 α，$\bar{\omega}$ 是否存在。我们在附录中的数字计算表明，同时满足式（7），（8）的解 α，$\bar{\omega}$ 是广泛存在的。

一般地，对于委托人来说，最优合同应该是在使得 U_j 取得最大值的 α 处实现，根据式（6），由于式（6）右端在 $[0,\bar{\alpha}]$ 上是 α 的连续函数，根据 weierstrass 定理，U_j 在 $[0,\bar{\alpha}]$ 上存在最大值。但是，我们目前只是证明了在约束 $\alpha^* < \alpha \leqslant \bar{\alpha}$ 以及代理人参与约束下，对于某些参数来说存在对于理性代理人情形下的委托-代理合同是帕累托改进的合同，当然这也保证了，对于某些参数来说存在对于理性代理人情形下的委托-代理合同是帕累托改进的最优合同。至于一般情形即在约束 $0 < \alpha \leqslant \bar{\alpha}$ 以及代理人参与约束下，是否仍然有同样的结论，有待于进一步研究。不过，如果此时最优合同是在 $0 < \alpha < \alpha^*$ 实现，则有 $\alpha < \dfrac{\beta}{b}$，公平均衡就不是“互惠”的而是“互损”的了，很难相信这种“互损均衡”合同会是理性代理人合同的帕累托改进。作为一个猜想，我们认为作为理性代理人合同的帕累托改进的非理性代理人合同一定在约束 $\alpha^* < \alpha \leqslant \bar{\alpha}$ 以及代理人参与约束下取得。这个猜想留给下一步工作去研究。

总之，我们已经获得的结果是，存在有关参数，使得委托人按照“动机公平”的“公平博弈”与非理性代理人签订“互惠性”委托-代理合同时，与 Holmstrom-Milgrom 模型里理性代理人假定下的最优委托-代理合同相比是帕累托改进的。存在最优的“互惠性”委托-代理合同，从而存在帕累托最优的委托-代理合同，并且最优的委托-代理合同是“互惠性”委托-代理合同。不难验证在 Holmstrom-Milgrom 模型里的均衡固定工资为 $\alpha_0=\bar{\omega}+\dfrac{b\rho\sigma^2-1}{2b(1+b\rho\sigma^2)}$，当有 $\alpha^*=-\dfrac{1}{\rho}\log\left(\dfrac{1+\mathrm{e}^{-\rho\bar{\alpha}}}{2}\right)>\bar{\omega}+\dfrac{b\rho\sigma^2-1}{2b(1+b\rho\sigma^2)}$时，就有 $\alpha>\alpha_0$，显然存在某些

参数并且在 $\bar{\alpha}$ 充分大时这个条件成立。$\alpha>\alpha_0$ 与 $a>\frac{\beta}{b}$ 意味着出现的是“互惠性”博弈均衡。即使 $\alpha^*=-\frac{1}{\rho}\log\left(\frac{1+e^{-\rho\bar{\alpha}}}{2}\right)>\bar{\omega}+\frac{b\rho\sigma^2-1}{2b(1+b\rho\sigma^2)}$ 不成立，只要 α 充分大（这要求 $\bar{\alpha}$ 充分大），这样的“互惠性”博弈均衡也是存在的，但是目前还不知道此时帕累托最优的委托-代理合约是不是“互惠性”博弈均衡。与蒲勇健（2007）的结果相比较，我们这里得到的结果之所以显得复杂一些的原因是，本文的分析考虑了代理人的物质效用效应，代理人不仅仅关注公平，也关注物质性的自利效用，他在两者之间存在着替换。

附录 数字模拟结果

由 $e^{-\rho\alpha}<\frac{1+e^{-\rho\bar{\alpha}}}{2}$ 可得，$\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}>\frac{1}{2}$，容易判断 $\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}$ 关于 α 单调增，再由 $0\leqslant\alpha^*\leqslant\alpha\leqslant\bar{\alpha}$，有 $\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\leqslant 1$，因此，$\frac{1}{2}<\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\leqslant 1$，所以，$-\frac{1}{2}\leqslant\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}<0$，进一步有 $0<\left|\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right|\leqslant\frac{1}{2}$。

a 要满足方程：

$$a=\frac{\beta}{b}-\frac{1}{A\rho b}\left[\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right]\exp\left\{\rho\left[\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right]\right\}, \tag{1}$$

则 a 是函数 $f(a)=\frac{\beta}{b}-\frac{1}{A\rho b}\left[\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right]\exp\left\{\rho\left[\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right]\right\}$ 的不动点，通过数值试验得到初步映象：当 $\frac{1}{A\rho b}$ 既不太小，又不太大时，$f(a)$ 有不动点，否则它在两个一大一小的数字间来回摆动，不能收敛。下面从分析角度考虑：

根据泛函分析知道，当 $f(a)$ 是压缩映射时，它有不动点。显然 $f(a)$ 是一个连续且有连续导数的函数，根据微分中值定理，对于 a_{n+1}、a_n，存在 a_{n+1} 和 a_n 之间的一个数 ξ，有

$$f(a_{n+1})-f(a_n)=-\frac{1}{A\rho b}\left[\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right]\exp\left\{\rho\left[\alpha+\beta\xi-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}\xi^2\right]\right\}\cdot(\beta-b\xi)(a_{n+1}-a_n).$$

由于 $\exp\left\{\rho\left[\alpha+\beta\xi-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}\xi^2\right]\right\}$ 在 $\xi=\frac{\beta}{b}$ 时取得极大值

$$\exp\left\{\rho\left[\alpha-\frac{1}{2}\rho\beta^2\sigma^2+\frac{b}{2}\beta^2\right]\right\}.$$

容易知道，通过方程（1）迭代得到的 a_{n+1}、a_n 都大于 0，从而 $\xi>0$，又 $b>0$，因此 $|\beta-b\xi|<\beta$，从而

$$|f(a_{n+1})-f(a_n)|<\frac{1}{2}\frac{\beta}{A\rho b}\exp\left\{\rho\left[\alpha-\frac{1}{2}\rho\beta^2\sigma^2+\frac{b}{2}\beta^2\right]\right\}|a_{n+1}-a_n|,$$

所以当$\frac{1}{2}\frac{\beta}{A\rho b}\exp\left\{\rho\left[\alpha-\frac{1}{2}\rho\beta^2\sigma^2+\frac{b}{2}\beta^2\right]\right\}<1$，即

$$\frac{\beta}{A\rho b}\exp\left\{\rho\left[\alpha-\frac{1}{2}\rho\beta^2\sigma^2+\frac{b}{2}\beta^2\right]\right\}<2 \tag{2}$$

时，这时，a 有解或存在（这仅仅是一个充分条件）。

下面考虑两个不等式，它们分别是正文中的式（7），（8）

$$\bar{\omega}-\alpha+\frac{(b\rho\sigma^2-1)}{b(1+b\rho\sigma^2)^2}-\frac{\sigma^2}{(1+b\rho\sigma^2)A}\left[\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right]$$

$$\cdot\exp\left\{\rho\left[\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right]\right\}>0, \tag{3}$$

$$\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\geqslant\bar{\omega}. \tag{4}$$

由（1）可得$-\frac{1}{A}\left[\frac{1}{2}-\frac{1-e^{-\rho\alpha}}{1-e^{-\rho\bar{\alpha}}}\right]\exp\left\{\rho\left[\alpha+\beta a-\frac{1}{2}\rho\beta^2\sigma^2-\frac{b}{2}a^2\right]\right\}=\rho ab-\rho\beta$，把它代入（2）式得$\bar{\omega}>\alpha-\frac{(b\rho\sigma^2-1)}{b(1+b\rho\sigma^2)^2}-\frac{\sigma^2}{(1+b\rho\sigma^2)}(\rho ab-\rho\beta)$，即

$$\bar{\omega}>\alpha-\frac{\beta^2(b\rho\sigma^2-1)}{b}-\rho ab\beta\sigma^2+\rho\beta^2\sigma^2. \tag{5}$$

令（5）的右边为 s1，（4）式的左边为 s2，则要证明存在满足条件的 α 和 $\bar{\omega}$，只需证明存在满足条件的 s1 和 s2，使得 s2$\geqslant\bar{\omega}>$s1，当然，就只需 s2$>$s1。

由于参数较多，函数较复杂，直接分析太复杂，因此下面采用数值计算。

编程思路：

（1）给出一组满足大于 0 要求的已知常数 A，$\bar{\alpha}$，b，ρ，σ（程序中分别用 A，a1，b，p，d 表示）；

（2）算出 β（程序中用 b1 表示）；

（3）根据（6）算出 α^*（程序中用 a2 表示）；

（4）然后以一个较小的步长（程序中用 0.01）在区间 $[\alpha^*, \bar{\alpha}]$ 顺次取 α（程序中用 a2 表示）的值，对每个不同的 α，连同 A，$\bar{\alpha}$，b，ρ，σ，β 代入不等式（2），检查是否满足条件（程序中用 try1 表示（2）左端的值），若不满足，则检查下一个 α；若满足，运用方程（1）的右边进行迭代，求出相应的 a（程序中用 aa2 表示），然后求出相应的 s1 和 s2，比较是否有 s1$<$s2。

程序运行结果是：只要满足不等式（2），就一定有 s1$<$s2，这意味着存在满足所有条件的 α 和 ω，并且这样的 α 和 $\bar{\omega}$ 还相当多。

注：在程序中，只要已知常数不满足不等式（2）（即 try1$>=$2），则意味着 a 不存在，就结束该次运算，开始计算下一个 α 的情况。

本程序是用 C＋＋编写的，可以在 VC＋＋6.0 上运行，读者可以改变常数试试。

$$\omega>\alpha-\frac{\beta^2(b\rho\sigma^2-1)}{b}-\rho ab\beta\sigma^2+\rho\beta^2\sigma^2. \tag{5}$$

令（5）的右边为 s1，（4）式的左边为 s2，则要证明存在满足条件的 α 和 β，只需证明存在满足条件的 s1 和 s2，使得 s2$\geqslant\omega>$s1，当然，就只需 s2$>$s1。

当 $A=2$、$\bar{\alpha}=2$、$b=2.5$；$\rho=1.5$、$\sigma=0.8$ 时，求得相应的 α 和 $\bar{\omega}$ 取值情况，见图1。图中阴影区域就是 $\bar{\omega}$ 的最小（即 s1）和最大取值（即 s2）范围。

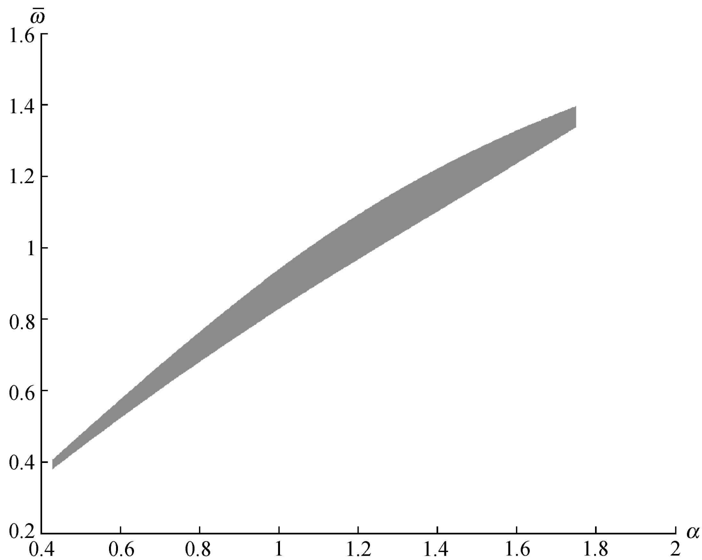

图1　$A=2$、$\bar{\alpha}=2$、$b=2.5$；$\rho=1.5$、$\sigma=0.8$

当 $A=1.5$、$\bar{\alpha}=1$、$b=1$；$\rho=2$、$\sigma=0.6$ 时，求得相应的 α 和 $\bar{\omega}$ 取值情况，见图2。图中阴影区域就是 $\bar{\omega}$ 的最小（即 s1）和最大取值（即 s2）范围。

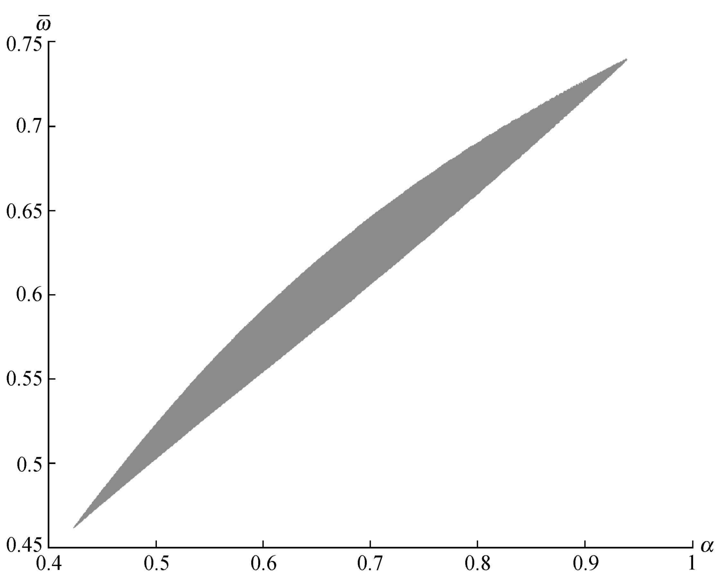

图2　$A=1.5$、$\bar{\alpha}=1$、$b=1$；$\rho=2$、$\sigma=0.6$

从图像看出，本问题应该有一般性的结论：只要给定参数满足不等式（2），就一定存在满足不等式的 α 和 $\bar{\omega}$。它们构成一个区域。

参 考 文 献

[1] Akerlof, G., “Labor Contracts as A Partial Gift Exchange”, *Quarterly Journal of Economics*, 1982, 97(4), 543—569.

[2] Akerlof G., and J. Yellen, “Fairness and unemployment”, *American Economic Review*, 1988, 78(2), 44—49.

[3] Fehr, E., G. Kirchsteiger, and A. Riedl, “Does Fairness Prevent Market Clearing: An Experimental Investigation”, *Quarterly Journal of Economics*, 1993, 108(2), 437—459.

[4] Holmstrom, B., “Moral Hazard and Observability”, *Bell Journal of Economics*, 1979, 10(1), 74—91.

[5] 拉丰, J.、D. 马赫蒂摩,《激励理论:委托-代理理论(第一卷)》,陈志俊等译。北京:中国人民大学出版社,2002 年。

[6] 蒲勇健,“植入‘公平’博弈的委托-代理模型:来自行为经济学的一个贡献”,《当代财经》,2007 年第 3 期,第 5—11 页。

[7] Mirrlees, J., “Notes on Welfare Economics, Information and Uncertainty”, in Balch M., D. McFadden, and S. Wu (eds.), *Essays on Economic Behavior under Uncertainty*. Amsterdam: North-Holland, 1974.

[8] Mirrlees, J., “The Theory of Moral Hazard and Unobservable Behavior: Part I”, Nuffield College, Oxford, Mimeo, 1975.

[9] Mirrlees, J., “The Optimal Structure of Authority and Incentives Within an Organization”, *Bell Journal of Economics*, 1976, 7(1), 105—131.

[10] Rabin, M., “Incorporating Fairness into Game Theory and Economics”, *American Economics Review*, 1993, 83(5), 1291—1302.

[11] Rabin, M., “A Perspective on Psychology and Economics”, European Economic Review, 2002, 46(4—5), 657—685.

[12] Rabin, M., “Review of Arrow, K., Colombatto, E., Perlman, M. and Schmidt, C. (eds.), *The Rational Foundations of Economic Behaviour*, Macmillan Press Ltd, 1996”, *Journal of Economic Literature*, 1997, 35(4), 2045—2046.

[13] Rabin, M., “Incorporating Behavioral Assumptions into Game Theory”, in Friedman, J. (ed.), *Problems of Coordination in Economic Activity*. Norwell, MA: Kluwer Academic Publishers, 1994.

[14] Rabin, M., “Fairness in Repeated Games”, Department of Economics, University of California at Berkeley, Working Paper No. 97-252, 1997.

[15] 张维迎,《博弈论与信息经济学》。上海:上海三联书店,1996 年。

A Principal-Agent Model with Fairness

YONGJIAN PU
(*Chongqing University*)

Abstract This paper develops a new principal-agent model in which the agent behaves both on rationality and a notion of fairness centered at reciprocity. A distinct feature of the model is that under many circumstances, the optimal contract provides a higher profit than what would be obtained when the agent behaves only on rationality.

JEL Classification D00, D82, C70