

# Economic incentives and social preferences: substitutes or complements?

Samuel Bowles & Sandra Polanía-Reyes

Journal of Economic Literature, 2012

## 1 主要结论

- 实验室实验和田野实验都证明了社会偏好对经济行为有重大影响。
- 不完全契约在理论和实证上都是很重要的。
- 激励通过影响信念和偏好而造成意想不到的影响。
- 激励有时是适得其反的 (counterproductive)。

## 2 概念

- 经济激励 (economic incentives): 试图通过改变经济成本或收益来改变行为。
- 社会偏好 (social preferences): 指利他主义、互惠互利、帮助他人的内在愉悦、不平等厌恶、道德承诺和其他促使人们帮助他人的动机, 而不是个人最大化个人利益的动机。
- 可分性假设 (separability assumption): 以前的经典假设认为行为函数中, 激励和社会偏好之间是独立的 (比如某人对公共物品补贴的反应与他之前的社会偏好无关)。
- 激励挤出 (crowd out) 社会偏好: 激励削弱了社会偏好, 使激励对行为的影响被削弱。
- 激励挤入 (crowd in) 社会偏好: 激励加强了社会偏好, 使激励对行为的影响被加强。

基于物质回报最大化的激励设计可能会适得其反, 这一主题在政治科学 (Taylor 1987; Grant 2011)、心理学 (Deci 1975)、社会学 (Healy 2006) 等领域多有研究, 但经济学角度的研究不太多, 因为以前基于经典假设, 社会偏好这一传导途径被简化忽略了。

人们常常对激励的存在而不是其程度作出反应, 尽管有时程度也是有影响的。行为对决策的本质非常敏感, 规则反映了规则制定人的意图 (比如是为利润最大化还是为公平)、指定人认为受众对激励的反应 (比如是否勤恳工作) 以及受众要采取的行为 (要勤恳到什么程度), 而上述信息也会影响受众的非经济动机。

内生偏好 (endogenous preference) 和 (state-dependent preference) 的区别, 激励对内生偏好的影响可长期维持, 激励传递了信号, 环境变了偏好不变; 但后一种偏好中, 激励改变了偏好更新的路径, 即改变了环境, 偏好也变了。

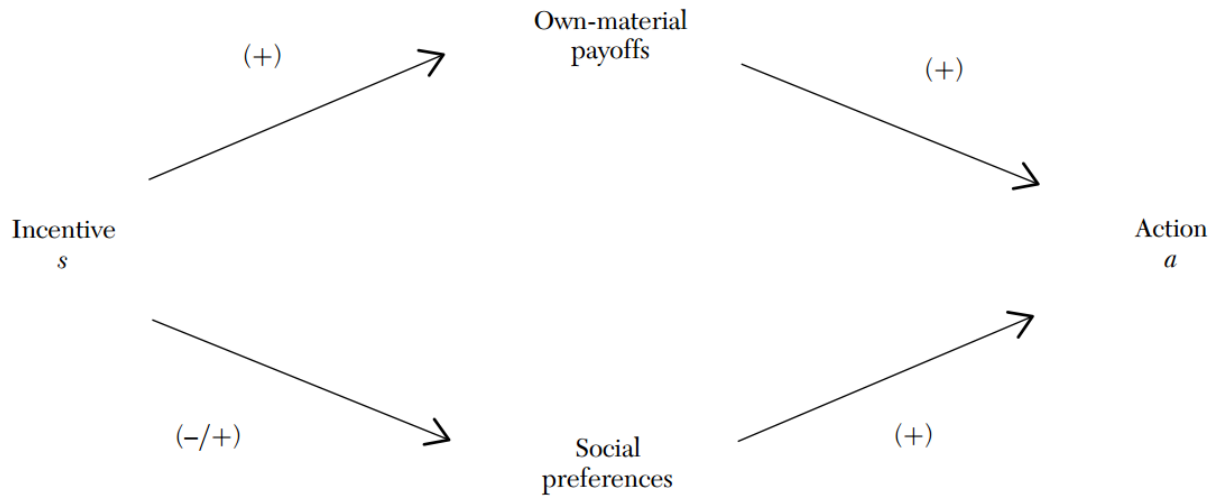


Figure 1. Crowding Effects of Incentives

激励怎样影响偏好：

- 为什么 state-dependent nature of preference 会导致挤出 (crowding out)，三个原因：“bad news”，“moral disengagement”，“control aversion”
- 通过改变对情形的主观感觉来改变社会偏好

Table 1: 经济激励对社会偏好的 endogenous effect 和 state-dependent effect

来源和特点	机制	描述
State-dependent preference: 激励影响了个体社会偏好的行为特点	Information: “bad news”	激励表达出激励设计者对受众和目标特征的信念, 并且可能显现了激励设计者对个人利益的不正当追求
	Framing: “moral disengagement”	激励显现了当下情况的类型以及受众应当做出的行为, 并且可能激发了自我回报最大化的思维模式
	Self-determination: “control aversion”	激励影响了受众对于自主性的感受, 激励表示着难以接受的控制
Endogenous preference: 激励影响了偏好形成的环境, 进而影响了偏好在人口中的分布	Conformist preference-updating	激励影响了人们获得偏好的过程, 这一影响能持续较长时间

### 3 Bad News: 激励提供了关于激励制定者的信息

**Trust Game:** 投资者 (investor) 可以向被信任者 (trustee) 转钱, 这笔钱会翻三倍, trustee 知道 investor 的选择, 可以反过来向 investor 转一些钱 (可以一分钱不转, 至多转出翻三倍后的所有钱)。

作者	被试	游戏	环境设定	结果	评论
Fehr and Rockenbach(2003)	德国学生 (238)	Trust Game	investor 向 trustee 转钱时, 可以设定一个期望回报, 如果 trustee 转回来的钱少于这个数目, 则 investor 可以向 trustee 收罚款, 这一选项事先向 trustee 声明。罚款数额固定。	当实行罚款时, trustee 转回来的钱变少, 可以罚款却不使用罚款使转回来的钱变多。	罚款体现了好意和信任。当投资者要求回报分配更多时, 罚款的负面效果越大。
Fehr and List (2004)	Costa Rican CEOs (126) 和学生 (76)	Trust Game	同上	CEO 比学生更愿意信任且更值得信任。罚款降低了收益。	
Borges and Irlenbusch (2007)	德国学生 (179)	Buyer-Seller Game		相比外生退货权 (比如法律规定), 如果卖家主动提供退货权, 买家会做出对卖家更有利的行为。	
Fehr and Schmidt (2007)	德国学生 (70)	Gift-Exchange Game	雇佣者可以选择提供两种合同: 1. 没有强制性措施的提成合同 2. 有罚款项的提成合同	大多数雇佣者不使用罚款, 纯提成合同的工资比另一个高 54%, 但利润区别不大。	被雇佣者认为不公平的雇佣者往往倾向于使用有罚款的合同, 并不太可能付提成。
Fehr et al. (2007)	德国学生 (130)	Gift-Exchange Game	三种合同: 1. 固定工资合同 2. 有罚款的合同 3. 纯提成合同。 不同的关系结构: 1. 雇佣关系 2. 买卖关系	纯提成合同下被试者贡献的比回报最大化纳什均衡还要多, 比罚款合同强。关系结构不影响结果。	提成合同带来了自主性。
Dickenson and Vileval (2008)	法国学生 (182)	Gift-Exchange Game with a computer task	陌生人或者伙伴关系。雇佣者的回报是否取决于被雇佣者的努力。	在伙伴关系中, 且雇佣者的回报取决于被雇佣者的努力时, 更少的监督带来更好的表现。	在伙伴关系中, 更多的监督传达出“制定规则者缺乏信任或者想损害被雇佣者来获利”的信息。
Irlenbusch and Ruchala (2008)	德国学生 (192)	Public Goods Game	外生的强制因素: 团队报酬, 给或者不给最高贡献者奖励。奖励高或低。个人奖励, 没有团队报酬。	高奖励提高了团队平均努力程度, 总剩余只在高奖励时提高, 但随时间衰减。纯团队报酬时, 参与者会比人人自利时贡献更多。纯个人奖励时的努力比人人自利水平低。	竞争结构减弱了自愿合作。

Ariely, Bracha, and Meier (2009)	美国学生 (161)	Charity giving based on task performance	外生强制因素：有或没有金钱报酬。捐款是公开的或者非公开的。“好”或“坏”的捐款原因。	公开场合下，被试为好原因捐款更多，在有激励下捐款减少。金钱激励提高了非公开场合捐款。	
Stanca, Bruni, and Corazzini (2009)	意大利学生 (96)	Gift-Exchange Game	在第一阶段行动时，玩家是否知道还有第二阶段行动。	在第一阶段玩家不知道有第二阶段行动时，第二阶段的行动者还回的额度与第一阶段玩家给出的额度更相关。	相比源于外部因素，在我方行为“发自内心”时，对方互惠动机更强。
Fehr and Gächter (2002b)	瑞士学生 (182)	Gift-Exchange Game	三种约束形式： 1. 固定工资合同（信任） 2. 含有罚款的合同 3. 提成激励合同	激励降低了被试的努力水平，罚款比提成的负面作用更大。激励降低了总剩余，提高了约束制定者的利润。	激励的影响源于制定约束者的公正、友善或者敌意。

#### 4 Moral Disengagement: 激励暗示了被允许的行为

与之前的实验不同，这里的激励都是外生的（实验者提供的），因此激励中不包含激励制定者的信念信息。

作者	被试	游戏	环境设定	结果	评论
Hoffman et al. (1994)	美国学生 (270)	Ultimatum Game; Dictator Game	角色是通过竞争得到的。有卖家买家结构，或者没有。双盲，或者没有。	角色竞争或者卖家买家结构下，开价更低，更少的开价被拒绝。开价者能准确估计对方的最低接受价位。独裁（dictator）游戏中，独裁者在双盲试验中开价更低。	制度设定的提示影响了行为：有产权（通过竞争获得）、有市场结构、纯匿名使人更加自利。
Schotter et al. (1996)	美国学生 (247)	Ultimatum Game	（Survival treatment）高回报的被试能存活到第2期（共两期）。（Non survival treatment）只有一期，开价者是随机决定的。同时行动或序贯行动。	存活的竞争压力使开价更低，对于低价的拒绝更少。	争取到的开价者地位、存活压力使开价者更自利。
Cardenas, Stranlund, and Willis (2000)	哥伦比亚林区居民 (112)	Common Pool Resource Game	外部约束：有一个低概率的检查和罚款。可以交流。	罚款导致了更多自利行为，在长期，减少过渡开采的效率降低。	轻微的外生强制罚款削弱了社会性动机行为。罚款使人从道德导向变为自利导向。
Cardenas (2004)	哥伦比亚乡下居民 (265)	Common Pool Resource Game	不同程度的外部约束（由低到高的罚款），声明了社会最优消耗，不能交流。可以交流，没有罚款和声明。	相比有罚款（高或低）没交流，在有交流（没罚款）的情况下更无私。高罚款对行为的影响比低罚款小6%。	边际挤出效应（“有没有”比“有多少”更重要）

Heyman and Ariely (2004)	美国学生 (150+90)	电脑任务和一个拼图任务	不同形式的报酬 (现金、糖果、以现金计价的糖果)。不同程度的金钱报酬 (无、低、中)。	糖果或者现金报酬下,报酬由低到中都使努力程度提升。在无报酬条件下,努力程度比低报酬 (现金或者现金计价糖果) 高, 但和糖果形式的报酬无差异。	报酬的程度和形式影响了表现, 金钱形式的报酬会作为金钱市场关系而非社会市场关系的强烈信号。金钱激励影响了动机, 市场交换的类型 (金钱还是社会) 影响了回报和动机的关系。
Bohnet and Baytelman (2007)	美国高管 (353)	Trust Game and a Dictator Game	没有交流, 玩之前面对面交流, 或者玩之后交流。外部约束 (游戏结束后是否有金钱惩罚)。陌生人或者伙伴。	重复和交流提高了转钱金额和报答金额。对低回报的惩罚降低了报答金额。	惩罚措施损害了内在信任, 降低了人们对信任的报答。
Houser et al. (2008)	美国学生 (532)	Gift-Exchange Game	有强制措施 (比如罚款合同)。惩罚是外生的或者是投资人选择的。	当对报答的要求越高 (罚款越高), 无论要求公平与否, 也无论惩罚是否外生, 惩罚降低了报答金额。	被试将惩罚理解为自利的价格或者自利的借口。
Mellstrom and Johansson (2008)	瑞典学生 (262)	提供被试体检机会, 为了献血	成为献血者, 是否提供金钱补偿。选择金钱补偿, 或者慈善捐赠相同金额。	激励降低了女性成为未来献血者的概率, 对男性没有影响。允许慈善捐赠消除了这种负面效应。	金钱激励使得社会偏好信号更难显现, 削弱了无私奉献的信号价值。慈善捐赠的选项弥补了这一点。
Li et al. (2009)	美国市民 (104)	Trust Game	以激励合同的形式选择惩罚 (比如如果报答金额少于预期就罚钱)	有惩罚威胁时, trustee 的报答减少。而且惩罚威胁减弱了关于社会回报的脑部活动, 加强了有关经济决策的脑部活动。	金钱惩罚促进了有关经济决策的神经活动, 同时减弱了有关社会回报的活动。
Henrich et al. (2010) and Barr et al. (2009) and personal communication from Barr and Henrich (March 2009)	15 个群体, 包括美国学生, 非洲工人, 亚马逊、北极、非洲的狩猎采集者 (428)	Dictator Game, Ultimatum Game, and Third-Party Punishment Game (TPG)	社会群体不同。被试保持角色不变 (主动或被动), 依次玩下列游戏: DG, UG, TPG。	对于 TPG 中的多数人, 惩罚激励不会提高开价, 反而有反作用。影响自利决策的因素 (财富, 收入, 住房面积) 是 TPG (但不是 DG) 中行动分布的显著因素。在 DG 中, 开价和宗教同盟 (伊斯兰教、天主教等) 有关, 但在 TPG 中无关。	TPG 中的惩罚减弱了道德思考 (来自于宗教等) 加强了基于经济需要的考虑。

## 5 Control Aversion: 激励可能影响内在动机和自主决定

人们有自主选择的动机，讨厌被控制。Principal Agent Game: Principal 可以选择是否设定最低产出，Agent 进行生产。

作者	被试	游戏	环境设定	结果	评论
Gneezy and Rustichini (2000b)	以色列学生 (160)	50 IQ 测试 + Principal Agent Game	对于 IQ 测试中选对答案有不同的金钱奖励 (很低, 低, 高, 没有)。	小奖励降低了表现, 大奖励提高了表现。	外在激励影响了内在动机。
Gneezy and Rustichini (2000b)	以色列学生 (180)	家庭捐款	对于捐款行为有不同的金钱奖励 (很低, 低, 高, 没有)。	没有奖励时表现最好, 其次是有高奖励, 最后是低奖励。	外在激励影响了内在动机。
Rustrom (2002)	美国学生 (110)	Creative task (汉诺塔游戏)	两种外在强制力 (奖励或惩罚)。不同程度 (没有, 弱, 强)。	没有奖励时提高最多, 其次是有高奖励, 最后是低奖励。惩罚降低了奖励。	激励仅在惩罚的情况下对表现有害。
Falk and Kosfeld (2006)	瑞士学生 (804)	Gift Exchange Game	外界设定了不同的表现下界 (低, 中, 高)。principal 可以选择是否实施下界限制。principal 可以选择 agent 的工资。	当 principal 有控制措施时, agent 往往依据下界表现。多数 principal 预期到了这一点, 实施了上下界限制, 提高了利润。	实施下界限制损害了接受者的自主性, 并发出了不信任的信号, 减弱了 agent 的互惠动机。
Xiao and Houser (2011)	美国学生 (72)	Public Goods Game	外生惩罚: 没有, 私下惩罚 (只有被惩罚的人知道惩罚的时间和数目), 公开惩罚 (大家都知道信息)	相比公开惩罚, 私下惩罚降低了贡献水平。	当激励是私下进行时, 激励挤出了合作, 但公开激励 (但是匿名的) 则提高了合作。
Gneezy (2003)	美国学生 (400)	Proposer-Responder Game	responder 有三种强制措施 (付出固定成本实施惩罚, 付出固定成本实施奖励, 无措施)。措施的程度不同 (强, 弱)。	惩罚或者奖励对表现的影响都不是单调的, 高强度措施下的 offer 最多, 低强度措施下的 offer 最少, 没有措施居中。	低强度的奖励或惩罚将行为模式从“道德的”变成“策略的”。

## 6 Crowding In

激励增强了行为, 并非削弱了行为。

Falk, Gächter, and Kovacs (1999)	匈牙利学生 (126,38)	Gift-Exchange Game	陌生人或者伙伴是否面对面 (社会压力)	伙伴关系提高了努力水平, 社会压力没有影响。	重复互动对于互惠以及公平分享的内在动机是很强的激励。
----------------------------------	----------------	--------------------	---------------------	------------------------	----------------------------

Gächter and Falk (2002)	奥地利学生 (116)	Gift-Exchange Game	陌生人或者伙伴	重复游戏提高了努力水平, 某些自利的玩家会策略性地互惠, 但在最后一轮选择最低努力	重复互动加强了互惠动机。
Tyran and Feld (2006)	瑞士学生 (102)	Public Goods Game	制裁的等级: 没有, 轻微, 严重。强制措施来自于外部还是内部产生 (投票)。	外部引入轻微制裁没有显著影响公共品的平均产出, 内部产生的制裁却有显著效果。	外部制裁提高了搭便车的成本, 但没有影响行为。可能的解释是公投产生的制裁包含了对搭便车的道德否定信号。
Herrmann et al. (2008a)	世界各地学生 (1120)	Public Goods Game (Partner)	金钱惩罚	惩罚环境下, 合作更强。	只有在强社会规范下惩罚才是对社会有利的, 只有这样, 惩罚才能引起羞愧的正面作用而不是怨恨的反面作用。
Rodriguez-Sickert, Guzmán, and Cárdenas (2008)	Rural Colombians from 5 communities (128)	Common Pool Resource Game	三种外部惩罚 (强制惩罚, 可以拒绝或者接受的惩罚, 无惩罚) 惩罚强度有高低之分	有惩罚时, 高合作水平。只有在惩罚是强制时合作水平才能维持高位, 如果玩家拒绝了惩罚, 合作水平会降低。	低贡献者的行为受到惩罚时, 其他人才得到满足, 否则其他人会以不合作的形式自行惩罚。
Carpenter et al. (2009)	美国志愿消防员 (217) 和非志愿消防员 (189)	Dictator Game	对志愿消防员的补偿 (是否给以及给多少)	补偿提高了不注重形象 (image concern) 的消防员的灭火表现, 但对注重形象的消防员没有影响。	
Gächter, Nosenzo, and Sefton (2010)	英国学生 (84)	Gift-Exchange Game with 3-member firms (one employer and two employees)	employee 序贯行动, employee1 知道工资差异的信息, employee2 由于在 1 之后行动而知道努力差异的信息。employer 可以给两个人高工资, 或者只给其中某一个高工资, 或者给两个人低工资。	相同的工资不影响努力水平, 尽管 employee2 的搭档比他的努力水平低。当搭档勤奋工作且工资与努力程度高度相关时, employee2 对 employer 的互惠更强。当工资发放不相等时, 将工资差异信息与努力差异信息分离, 不影响 employee 对 employer 的互惠动机	对于不同的人制定不同的工资, 削弱了互惠动机。employee 更会因为 employer 的赏识 (工资和努力高度相关) 而努力表现, 慷慨 (给所有人一样的高工资) 并不会激发这种动机。
Barr (2001)	津巴布韦村民 (602)	Public Goods Game	两种极端的非金钱惩罚 (1. 每个玩家向其余玩家公开自己的贡献 2. 每个人对他人公开评价, 表扬或批评)	在两个惩罚措施实施后, 贡献提升了。	非金钱惩罚引入了羞愧或者其他社会情绪 (自利个体的最佳反应是不理会它们)

## 7 Control Aversion: 激励可能影响内在动机和自主决定

人们有自主选择的动机，讨厌被控制。Principal Agent Game: Principal 可以选择是否设定最低产出，Agent 进行生产。

作者	被试	游戏	环境设定	结果	评论
Gneezy and Rustichini (2000b)	以色列学生 (160)	50 IQ 测试 + Principal Agent Game	对于 IQ 测试中选对答案有不同的金钱奖励（很低，低，高，没有）。	小奖励降低了表现，大奖励提高了表现。	外在激励影响了内在动机。
Gneezy and Rustichini (2000b)	以色列学生 (180)	家庭捐款	对于捐款行为有不同的金钱奖励（很低，低，高，没有）。	没有奖励时表现最好，其次是有高奖励，最后是低奖励。	外在激励影响了内在动机。
Rustrom (2002)	美国学生 (110)	Creative task (汉诺塔游戏)	两种外在强制力（奖励或惩罚）。不同程度（没有，弱，强）。	没有奖励时提高最多，其次是有高奖励，最后是低奖励。惩罚降低了奖励。	激励仅在惩罚的情况下对表现有害。
Falk and Kosfeld (2006)	瑞士学生 (804)	Gift Exchange Game	外界设定了不同的表现下界（低，中，高）。principal 可以选择是否实施下界限制。principal 可以选择 agent 的工资。	当 principal 有控制措施时，agent 往往依据下界表现。多数 principal 预期到了这一点，实施了上下限制，提高了利润。	实施下界限制损害了接受者的自主性，并发出了不信任的信号，减弱了 agent 的互惠动机。
Xiao and Houser (2011)	美国学生 (72)	Public Goods Game	外生惩罚：没有，私下惩罚（只有被惩罚的人知道惩罚的时间和数目），公开惩罚（大家都知道信息）	相比公开惩罚，私下惩罚降低了贡献水平。	当激励是私下进行时，激励挤出了合作，但公开激励（但是匿名的）则提高了合作。
Gneezy (2003)	美国学生 (400)	Proposer-Responder Game	responder 有三种强制措施（付出固定成本实施惩罚，付出固定成本实施奖励，无措施）。措施的程度不同（强，弱）。	惩罚或者奖励对表现的影响都不是单调的，高强度措施下的 offer 最多，低强度措施下的 offer 最少，没有措施居中。	低强度的奖励或惩罚将行为模式从“道德的”变成“策略的”。

## 8 Crowding In

激励增强了行为，并非削弱了行为。

Falk, Gächter, and Kovacs (1999)	匈牙利学生 (126,38)	Gift-Exchange Game	陌生人或者伙伴是否面对面（社会压力）	伙伴关系提高了努力水平，社会压力没有影响。	重复互动对于互惠以及公平分享的内在动机是很强的激励。
----------------------------------	----------------	--------------------	--------------------	-----------------------	----------------------------



Gächter and Falk (2002)	奥地利学生 (116)	Gift-Exchange Game	陌生人或者伙伴	重复游戏提高了努力水平, 某些自利的玩家会策略性地互惠, 但在最后一轮选择最低努力	重复互动加强了互惠动机。
Tyran and Feld (2006)	瑞士学生 (102)	Public Goods Game	制裁的等级: 没有, 轻微, 严重。强制措施来自于外部还是内部产生 (投票)。	外部引入轻微制裁没有显著影响公共品的平均产出, 内部产生的制裁却有显著效果。	外部制裁提高了搭便车的成本, 但没有影响行为。可能的解释是公投产生的制裁包含了对搭便车的道德否定信号。
Herrmann et al. (2008a)	世界各地学生 (1120)	Public Goods Game (Partner)	金钱惩罚	惩罚环境下, 合作更强。	只有在强社会规范下惩罚才是对社会有利的, 只有这样, 惩罚才能引起羞愧的正面作用而不是怨恨的反面作用。
Rodriguez-Sickert, Guzmán, and Cárdenas (2008)	Rural Colombians from 5 communities (128)	Common Pool Resource Game	三种外部惩罚 (强制惩罚, 可以拒绝或者接受的惩罚, 无惩罚) 惩罚强度有高低之分	有惩罚时, 高合作水平。只有在惩罚是强制时合作水平才能维持高位, 如果玩家拒绝了惩罚, 合作水平会降低。	低贡献者的行为受到惩罚时, 其他人才得到满足, 否则其他人会以不合作的形式自行惩罚。
Carpenter et al. (2009)	美国志愿消防员 (217) 和非志愿消防员 (189)	Dictator Game	对志愿消防员的补偿 (是否给以及给多少)	补偿提高了不注重形象 (image concern) 的消防员的灭火表现, 但对注重形象的消防员没有影响。	
Gächter, Nosenzo, and Sefton (2010)	英国学生 (84)	Gift-Exchange Game with 3-member firms (one employer and two employees)	employee 序贯行动, employee1 知道工资差异的信息, employee2 由于在 1 之后行动而知道努力差异的信息。employer 可以给两个人高工资, 或者只给其中某一个高工资, 或者给两个人低工资。	相同的工资不影响努力水平, 尽管 employee2 的搭档比他的努力水平低。当搭档勤奋工作且工资与努力程度高度相关时, employee2 对 employer 的互惠更强。当工资发放不相等时, 将工资差异信息与努力差异信息分离, 不影响 employee 对 employer 的互惠动机	对于不同的人制定不同的工资, 削弱了互惠动机。employee 更会因为 employer 的赏识 (工资和努力高度相关) 而努力表现, 慷慨 (给所有人一样的高工资) 并不会激发这种动机。
Barr (2001)	津巴布韦村民 (602)	Public Goods Game	两种极端的非金钱惩罚 (1. 每个玩家向其余玩家公开自己的贡献 2. 每个人对他人公开评价, 表扬或批评)	在两个惩罚措施实施后, 贡献提升了。	非金钱惩罚引入了羞愧或者其他社会情绪 (自利个体的最佳反应是不理会它们)