

亲社会努力影响奖赏加工的双阶段模型*

郑亚 杨子阳

(广州大学教育学院奖赏与社会认知研究中心, 广州 510006)

摘要 亲社会行为往往需要个人付出成本为他人带来收益, 当前研究大多关注他人收益在亲社会行为中的作用, 忽略了自我成本的影响。少量关注自我成本的研究聚焦努力成本与他人奖赏的决策机制, 忽视了自我努力调控他人收益的完整神经动态变化过程。在前期研究中, 我们发现利己努力既会前瞻性折扣奖赏价值也会回溯性增加奖赏价值的悖论机制。我们推测这种努力与奖赏的双阶段模型可能是亲社会行为的一个新机制。为证实这一假说, 我们将结合行为实验与脑电和脑成像等技术, 比较个体付出认知努力为自己和为他人赢得奖赏的神经动态变化, 并考察这种神经动态过程在不同努力类型和收益领域的一般性和特异性机制, 从而阐明亲社会努力影响奖赏加工的认知神经机制。该机制的阐明有助于弥补主流理论的不足, 为理解亲社会行为中自我成本和他人收益之间的关系及促进亲社会行为提供新视角。

关键词 亲社会行为; 认知努力; 奖赏; 努力悖论

分类号 C91

1 研究意义

亲社会行为在人类社会中普遍存在, 对于维护人际关系、保持团体稳定以及增强社会凝聚力具有重要作用。研究表明, 亲社会倾向能正向预测身体和心理健康水平, 并有效降低人类的死亡率(Brown & Brown, 2015)。因此, 阐明亲社会行为的心理过程和认知神经机制是心理学和认知神经科学领域的前沿热点问题之一(谢晓非, 寇彧, 2024)。

大多数亲社会行为需要个体付出努力为他人带来收益。小到为老人让座、给路人指路, 大到参与志愿服务、抗震救灾, 都需要个体付出努力并关注他人的收益。现有关于亲社会行为的理论和研究主要关注他人收益, 仅有少量研究聚焦于自我努力成本在亲社会行为中的作用。这些亲社会努力的研究探讨了努力成本与奖赏收益的决策机制, 发现个体在决定为他人(相较于为自己)付出努力时表现出更强的奖赏折扣效应(Forbes et al., 2024; Lockwood et al., 2022)。然而, 付出努力获得奖赏收益是一个多阶段的动态认知过程, 至少包括努力执行前的价值决策和执行后的价值评估(Husain & Roiser, 2018; Ruff & Fehr, 2014)。现有关于亲社会努力的研究仅限于探讨努力成本在价值决策阶段对他人收益的影响, 尚未揭示努力执行如何影响后续对他人收益的评估, 导致努力成本调控他人收益的神经动态变化过程仍不明确。基于此, 本研究首次提出亲社会努力影响奖赏加工的双阶段模型, 通过对比为自我获益和为他

收稿日期: 2026-3-13

*国家自然科学基金项目(32571255)资助。

通信作者: 郑亚, E-mail: zhengya1982@gmail.com

ChinaXiv:202606.00002v1

人获益时，努力执行之前和努力执行之后的奖赏价值的神经动态过程（研究一），以及考察这种神经动态过程在不同努力类型和收益领域的领域一般性与特异性机制（研究二），着重探讨亲社会努力在执行前后影响奖赏主观价值的认知神经机制。该机制的阐明可以弥补主流理论过分强调他人收益的不足，揭示亲社会行为中努力成本和奖赏收益的动态整合过程，帮助研究者理解和预测亲社会行为的产生和异常、采取针对性干预措施，从而提升个体的亲社会动机，促进积极友善的人际关系，增进社会和谐，助力践行社会主义核心价值观。

2 国内外研究进展与述评

2.1 亲社会行为的成本-收益框架

亲社会行为作为一个广泛的行为类别，既包括分享、安慰和合作等行为，也包括捐赠物品、金钱和提供志愿服务等行为。在本研究中，我们将亲社会行为定义为以个人成本为代价为他人带来收益的行为(Pfattheicher et al., 2022)。近年来，越来越多的研究者开始倡导从成本-收益权衡框架来理解亲社会行为(Contreras-Huerta et al., 2020)。该定义不仅强调了对他人收益的关注，同时也突显了个人成本的重要性。与既往理论和研究强调他人收益、忽视自我成本不同，成本-收益框架认为，亲社会行为是个人成本和他人收益权衡的结果。个体的亲社会行为既有可能是因为其对于他人收益的关注，也有可能是源自于其对自我成本的不敏感。利用成本-收益框架理解亲社会行为具有独特优势。首先，成本-收益框架可以解释多种亲社会行为。该框架中的成本包括努力、时间、金钱或压力等，收益则包括赢得金钱、避免伤害、提供食物或情绪支持等，因而囊括了多种亲社会行为，具有更大的解释力度。其次，成本-收益框架允许在同一实验范式中独立操控和测量自我成本和他人收益，有助于回答亲社会行为是源自于对他人收益的过分关注，还是对个人成本的不敏感，或是二者兼而有之。

值得注意的是，成本-收益框架中的个人成本并非单一维度，而是包含多种不同形式。研究中常见的个人成本包括金钱成本、时间成本、机会成本、道德成本、疼痛成本以及努力成本等多种形式。不同类型的成本在可量化程度、主观厌恶性、情境依赖性及其所涉及的心理与神经机制上存在差异。具体而言，金钱成本因其具有较高的可量化性、便于在实验中精确操控等特点而最早被引入相关领域的研究，但其影响也容易受到个体社会经济地位和社会文化等因素的调节(Vanags et al., 2025)；时间成本和机会成本则更多体现为个体在帮助他人时放弃其他活动或潜在收益的代价，往往具有较强的情境依赖性(Persson et al., 2024)；道德成本和疼痛成本则分别涉及规范违背和身体伤害所带来的负性体验(Lockwood et al., 2025; Qu et al., 2020)；而努力成本主要表现为个体完成特定任务时在体力或脑力上的消耗程度

(Contreras-Huerta et al., 2020)。从研究范式来看, 现有基于成本-收益框架的亲社会行为研究主要集中于金钱、疼痛和道德成本。关于金钱成本的研究多采用经济博弈范式, 例如考察个体是否愿意向慈善机构捐赠金钱(Engel, 2011)或付出金钱让他人获益(Sellitto et al., 2021), 这些研究普遍发现, 个体在此类情境中倾向于优先考虑自身经济利益, 呈现出以利己为主导的行为模式。相比之下, 采用道德成本的研究则揭示了明显的亲社会倾向。当个体需要在经济获利与施加痛苦(如电击)之间进行权衡时, 相比于电击自身, 个体对电击他人表现出更强的厌恶, 需要更高的金钱补偿才愿意增加他人所承受的电击次数, 表现出超利他倾向(Crockett et al., 2014; Crockett et al., 2017), 甚至愿意直接替他人分担身体疼痛(Hein et al., 2010)。然而, 相较于这些成本形式, 努力成本(包括认知和体力努力)是亲社会行为中更为常见的成本形式。例如, 为陌生人按住电梯、帮同事评阅项目书、向慈善机构提供志愿服务等亲社会行为都需要我们付出不同程度的认知或体力努力。因此, 理解努力成本在亲社会行为中的作用及其相关认知神经机制具有更大的生态效度和现实意义。

2.2 努力与奖赏之间存在悖论关系

作为一种令人厌恶的成本, 努力会在执行之前前瞻性降低奖赏价值。人们对未来奖赏的估值会随所需努力水平的提升而降低, 表现出努力折扣现象(Kool & Botvinick, 2018)。内部成本理论认为, 努力产生了主观成本, 该成本源自于有机体内部资源、工作记忆或表征能力的有限性(Shenhav et al., 2017); 机会成本理论认为, 选择执行高努力任务意味着失去了完成其他任务以获得奖赏的机会, 表现为一种机会成本(Kurzban et al., 2013)。支持努力贬值的证据包括参与负性情绪编码的前脑岛和前扣带回的激活随着预期努力水平的增加而增加(Suzuki et al., 2021), 而以腹侧纹状体为代表的奖赏回路编码了预期奖赏值减去预期努力值的净值(Westbrook et al., 2020)。同时, 努力也会在执行之后回溯性增加奖赏价值。例如, 人们愿意支付更高的价格去购买需要自己组装的家具(Norton et al., 2012), 倾向于选择更费力的任务而非无所事事(Wu et al., 2023), 并能从付出努力的过程中获得价值感(Clay et al., 2022)。努力合理化理论认为, 个体倾向于对花费更多努力获得的结果赋予更高的主观价值, 从而合理化先前付出的努力(Aronson & Mills, 1959); 心理对比理论认为, 与低努力任务相比, 高努力任务诱发的厌恶状态与后续奖赏结果形成了更为强烈的对比, 从而增加了奖赏的主观价值(Zentall, 2010)。支持努力增值的证据包括认知努力增强了前扣带回、伏隔核和前脑岛对奖惩结果的敏感性(Hernandez Lallement et al., 2014)以及腹侧纹状体对行为反馈的神经反应(Dobryakova et al., 2017)。

这种努力既折扣也增加奖赏价值的努力悖论现象主要源自于不同的认知加工阶段: 努力

贬值主要发生在努力执行之前的决策和线索评估阶段,而努力增值主要发生在努力完成之后的奖赏评估阶段(Inzlicht et al., 2018)。凭借其优秀的时间分辨率, EEG 技术能够区分时间上非常接近的认知过程,特别适用于研究努力和奖赏的神经动态关系。一方面,认知努力降低了线索评估阶段的线索 P3 和关联性负变(CNV)等成分的幅值(Zhang & Zheng, 2022; Zheng et al., 2024)。另一方面,认知努力提升了后续反馈评估诱发的 EEG 指标的敏感性,包括 RewP(Bogdanov et al., 2022; Yi et al., 2020)、反馈 P3(Lopez-Gamundi et al., 2024; Yi et al., 2020)和 Delta 功率(Bogdanov et al., 2022)。采用体力努力任务的 EEG 研究也获得了类似结论:体力努力增强了 RewP(Pan et al., 2023; Umemoto et al., 2022)和晚期正电位(LPP; Wu & Zheng, 2023)的幅值。

然而,上述关于努力折扣与努力增值的理论解释及实证研究主要来自非社会决策领域,其能否推广至亲社会情境尚不清楚。在非社会决策中,个体权衡的是自身所获奖赏与所付努力之间的关系;而在亲社会情景中,努力所换取的结果由他人获得,个体的价值计算不仅需要整合努力成本,还需要对他人收益赋予主观社会价值(Ruff & Fehr, 2014),即从追求自我收益最大化转向为他人付出努力以帮助其获益的社会目标(Lockwood et al., 2020)。基于这些考虑,利己和利他努力在努力与奖赏的关系上可能存在以下两方面的差异。第一,从理论预测来看,不同理论框架对利己努力与利他努力的预测存在分离。就前瞻性努力贬值而言,内部成本假设对两类努力的预测相似,因为无论受益对象是自己还是他人,努力本身产生的内部成本是类似的。然而,机会成本假说的预测可能不同:为他人付出努力时,个体放弃了将同等努力用于自身收益的机会,因此利他努力会伴随更大的机会成本。就回溯性努力增值而言,两种理论的预测同样存在差异。努力合理化理论强调个体对自身付出的意义建构与认知合理化,是一种晚期的、自上而下的加工过程;当受益对象为他人时,个体缺乏直接的自我相关动机来合理化付出,因此利他努力在回溯性阶段可能不会产生增值效应,甚至可能出现努力贬值。相反,心理对比理论描述的是一种自动的、自下而上的加工过程,这种努力与奖赏之间的对比可能在价值评估的早期阶段以不区分受益对象的方式进行,从而利己与利他努力产生相似的增值效应。第二,从边界条件来看,亲社会努力涉及到对他人收益赋予主观社会价值,因此可能表现出更大的情境敏感性,更容易受到社会因素的调节。首先,社会距离是影响亲社会行为的重要因素,人们通常对与自己的群际关系(内外群体)或心理距离(亲疏关系)更近的他人表现出更多的亲社会行为(Jones, 2021)。其次,个体层面的共情能力和互惠影响也可能调节亲社会努力的价值评价。共情-利他假说认为,对他人困境的共情会诱发减轻或消除其困境的动机,从而促进利他行为(Batson et al., 1991);而互惠理论则指出,善意

行为会通过社会规范等机制诱发受助者的回馈行为(Gouldner, 1960)。共情水平较高或互惠预期较强的个体,更可能倾向于为他人努力,并由此将获得的奖赏赋予更高的主观价值。综上,亲社会情境下的努力-奖赏之间的关系可能受到社会距离、共情能力和互惠信念等多种社会因素的调节,表现出比非社会情境更大的可塑性。

2.3 亲社会努力研究进展

目前仅有少量研究考察了亲社会努力的认知神经机制。Lockwood 等人(2017)首次将基于努力的奖赏决策范式运用到亲社会研究,要求被试在一个需要付出体力努力获得高奖赏的选项和一个休息获得低奖赏的选项之间进行决策。结果发现,当努力水平较低时,被试为自己和匿名他人付出努力赢得奖赏的意愿相同,而当努力水平较高时,被试更多地选择休息选项而非为他人付出努力,表现出亲社会淡漠(prosocial apathy)。计算建模分析发现,这种亲社会淡漠源自于体力努力以抛物线方式折扣了奖赏价值,导致他人奖赏在高努力条件下产生了更大的贬值效应。在神经机制方面,为他人进行努力决策特异性激活了前扣带回,其激活程度与个体的共情水平和努力投入程度存在正相关。相反,为自己进行努力决策特异性激活了腹侧被盖区和腹侧脑岛(Lockwood et al., 2022)。另有研究发现,与为自己赢得奖赏相比,个体更愿意为慈善机构付出体力努力以赢得奖赏(Imas, 2014)。最近的一项研究在认知努力中也发现了亲社会淡漠现象:个体的努力回避意愿在自己、慈善机构和匿名他人条件下依次增强(Depow et al., 2022)。

亲社会努力受多种因素的影响。与年轻人相比,老年人表现出了更强的亲社会努力倾向,更愿意为匿名他人付出体力努力赢得奖赏(Lockwood et al., 2021)。相较于正常青少年,具有品行障碍的青少年更不愿意付出体力努力为他人赢得奖赏,而那些具有高冷酷无情特质的品行障碍青少年在执行努力时投入的努力程度更低(Gaule et al., 2024)。另外,亲社会努力意愿与个体的共情关注和观点采择能力存在正相关,与情感淡漠、心理病理和述情障碍等存在负相关(Contreras-Huerta et al., 2022)。个体为他人付出体力努力赢得奖赏的意愿在经历急性应激后降低,并伴随背侧前扣带回和前脑岛活动的削弱(Forbes et al., 2024),而多巴胺缺乏的帕金森病患者在服用多巴胺药物后表现出了更强的亲社会努力意愿(Talbot et al., 2025)。

3 问题提出

总体而言,目前关于亲社会努力的研究刚刚起步,现有研究通过比较利他和利己努力的决策机制,突显了努力成本在亲社会行为中的关键作用,发现了亲社会努力受到年龄、急性应激、共情特质等因素的影响,揭示了亲社会努力与自我努力不同的认知和神经基础。然而,

现有研究存在以下主要问题：第一，现有研究忽视了亲社会努力和奖赏加工之间的动态整合过程。通过文献进行梳理，我们发现利己努力对奖赏加工的影响呈现悖论模式：在努力付出之前，奖赏估值随着努力水平增加表现出折扣现象(Zhang & Zheng, 2022; Zheng et al., 2024)，而在努力付出之后，奖赏估值随着努力水平增加表现出增值现象(Bogdanov et al., 2022; Lopez-Gamundi et al., 2024; Wu & Zheng, 2023)。目前关于亲社会努力的研究主要聚焦于努力付出之前的决策阶段，而对努力完成之后的奖赏评估阶段关注不足，使得在整合统一框架下研究亲社会努力缺乏完整的数据支持，从而无法全面理解亲社会行为中努力成本和奖赏收益之间的动态关系。基于机会成本理论(Kurzban et al., 2013)和努力合理化理论(Aronson & Mills, 1959)，我们预期亲社会努力与利己努力可能存在不同的认知神经机制。不同于利己努力的先贬值后增值的努力悖论模式，对于陌生他人的亲社会努力可能存在努力执行前后持续性折扣奖赏价值的双重贬值模式。

第二，现有研究尚未探讨亲社会努力影响奖赏加工的领域一般性/特异性机制。从成本-收益的视角来看，亲社会行为由自我努力成本和他人收益共同决定。其中，努力成本主要包括体力努力和认知努力两种基本形式，而他人收益既会出现在帮助别人获得奖赏或避免损失的经济博弈领域，也会发生在帮助别人避免伤害的道德决策领域。由此引发的一个基本问题是，亲社会努力影响奖赏加工的神经动态整合过程是遵循领域一般性原则还是领域特异性原则？现有亲社会努力研究在努力成本方面主要关注体力努力，忽视了认知努力；在收益方面聚焦于为他人赢得奖赏，忽视了其他收益领域。这种聚焦于单一努力成本和收益领域的研究无法回答亲社会努力影响奖赏加工的领域一般性/特异性问题。因此，有必要从努力类型和收益领域两个维度系统考察亲社会努力对奖赏加工的影响机制。来自非社会领域的研究表明，体力和认知努力存在既重叠又分离的行为和神经机制(Bustamante et al., 2023; Chong et al., 2017; Jiang & Zheng, 2024)。另外，个体的亲社会性会受到收益领域的影响：相较于赢钱领域，人们在损失或伤害领域表现出更多的亲社会行为，甚至愿意牺牲自我利益以避免对他人造成损失(Sellitto et al., 2021)或伤害(Liu et al., 2020)。这些研究表明亲社会努力和奖赏的动态关系在不同收益领域呈现不同模式。总之，亲社会努力对奖赏加工的影响机制如何受到成本类型（体力努力、认知努力）和收益领域（奖赏获得、损失回避、伤害回避）的调控仍不明确。

4 研究构想

针对上述问题，本研究提出亲社会努力影响奖赏加工的双阶段模型（图1）。基于该模

型，我们预期亲社会努力对奖赏加工的影响存在两种主要动态关系。当受益对象为自己时，努力对奖赏评估表现出努力执行前贬值而努力执行后增值的悖论模式；而当受益对象为匿名陌生他人时，亲社会努力对奖赏评估持续表现出折扣效应，形成双重努力贬值模式。同时，亲社会努力影响奖赏加工的动态关系在不同努力成本和收益领域之间表现出领域一般性或特异性特点。

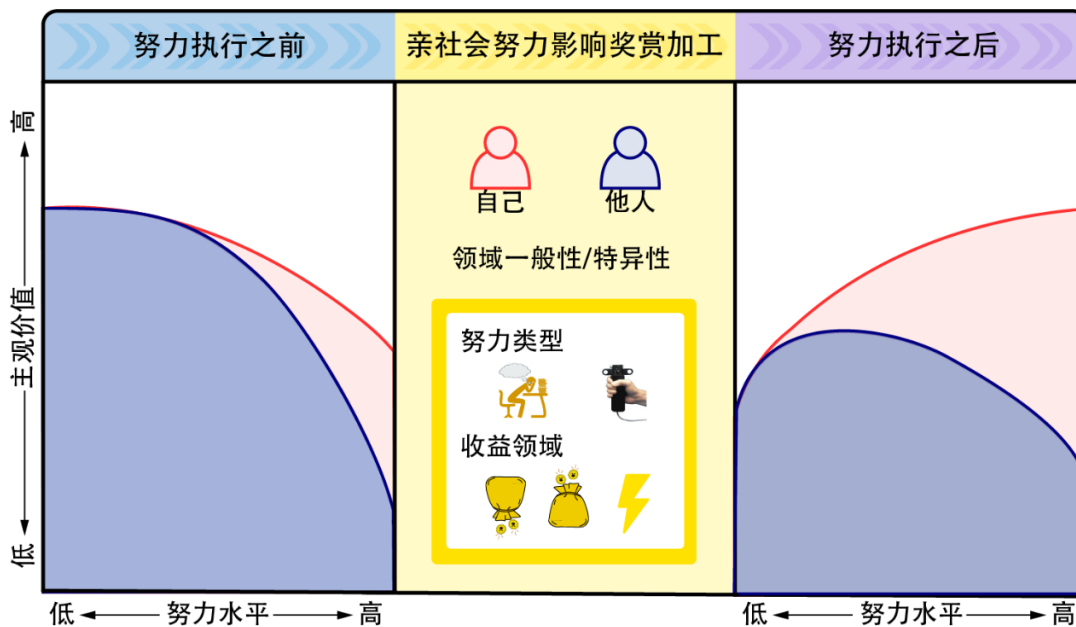


图1 亲社会努力影响奖赏加工的双阶段模型图。该模型的核心假设聚焦于努力在执行之前和之后如何影响奖赏的主观价值。红色实线代表自我努力影响自我奖赏的主观价值（先贬值后增值），蓝色实线代表自我努力影响他人奖赏的主观价值（双重贬值）。

主观价值包括奖赏刺激在努力执行前后诱发的主观报告、行为偏好、EEG或fMRI反应等。

基于努力与奖赏加工的双阶段模型，我们计划围绕“亲社会努力影响奖赏加工的认知神经机制”开展研究工作。我们聚焦于当前研究较少的认知努力，以脑电技术为主，结合行为实验和脑成像技术，对比为自我获益和为他人获益时，努力执行之前和努力执行之后的奖赏价值的神经动态过程（研究一），同时考察这种神经动态过程在不同努力类型和收益领域的领域一般性与特异性机制（研究二），以全面理解亲社会努力对奖赏加工的作用机制。具体研究内容分述如下：

4.1 研究一：亲社会努力影响奖赏加工的基本机制

研究一通过三个实验（实验1、实验2a、实验2b），分别从行为现象、神经动态过程和神经基础考察亲社会努力影响奖赏加工的基本机制。在实验1中，我们对比被试为自己和匿名陌生他人付出努力赢得奖赏时，在努力执行之前的奖赏决策倾向和努力执行之后获得奖赏诱发的情绪反应。其中，实验1主要揭示基本的认知机制，实验2进一步结合EEG和fMRI技术，考察相关神经动态过程及其神经基础。

4.1.1 实验1 亲社会努力影响奖赏加工的认知机制：行为研究

实验 1 旨在从行为层面考察亲社会努力影响奖赏加工的认知机制。被试依次完成角色分配任务（图 2A）、亲社会努力任务（图 2B）和决策任务（图 2C），以对比被试为自己和匿名陌生他人付出努力赢得奖赏时，在努力执行之前的奖赏决策倾向和努力执行之后获得奖赏诱发的情绪反应。在亲社会努力任务中，被试为自己和匿名陌生他人付出认知努力赢得奖赏，并以该奖赏所诱发的愉悦程度为行为指标进行比较。而在决策任务中，被试需要在一个基线选项和一个高努力选项之间进行选择，比较其为自己和为他人付出认知努力以获得奖赏时的决策倾向，并利用计算建模的方法估算努力折扣程度。为控制其他社会因素的干扰，研究将通过匿名化程序对“他人”身份进行处理，使被试在明确他人真实存在的同时，尽可能降低社会互动、熟悉性和互惠预期等因素的影响。认知努力任务将采用经典的注意加工范式（快速序列视觉呈现范式）操纵不同努力水平，从而比较在不同受益对象条件下，努力水平如何影响个体的奖赏决策与奖赏体验。我们预期亲社会努力对奖赏加工的影响在努力执行前后表现出双重贬值模式。具体而言，在努力执行前的决策阶段，与为自己努力相比，被试更不愿意为他人付出努力获得奖赏，其努力折扣程度（ k 值）更高。在努力执行后的评价阶段，被试对为自己获得的奖赏愉悦度评价随着努力水平的提升而升高，表现出努力增值现象，对为他人获得的奖赏愉悦度评价随着努力水平的提升而降低，表现出努力折扣现象。

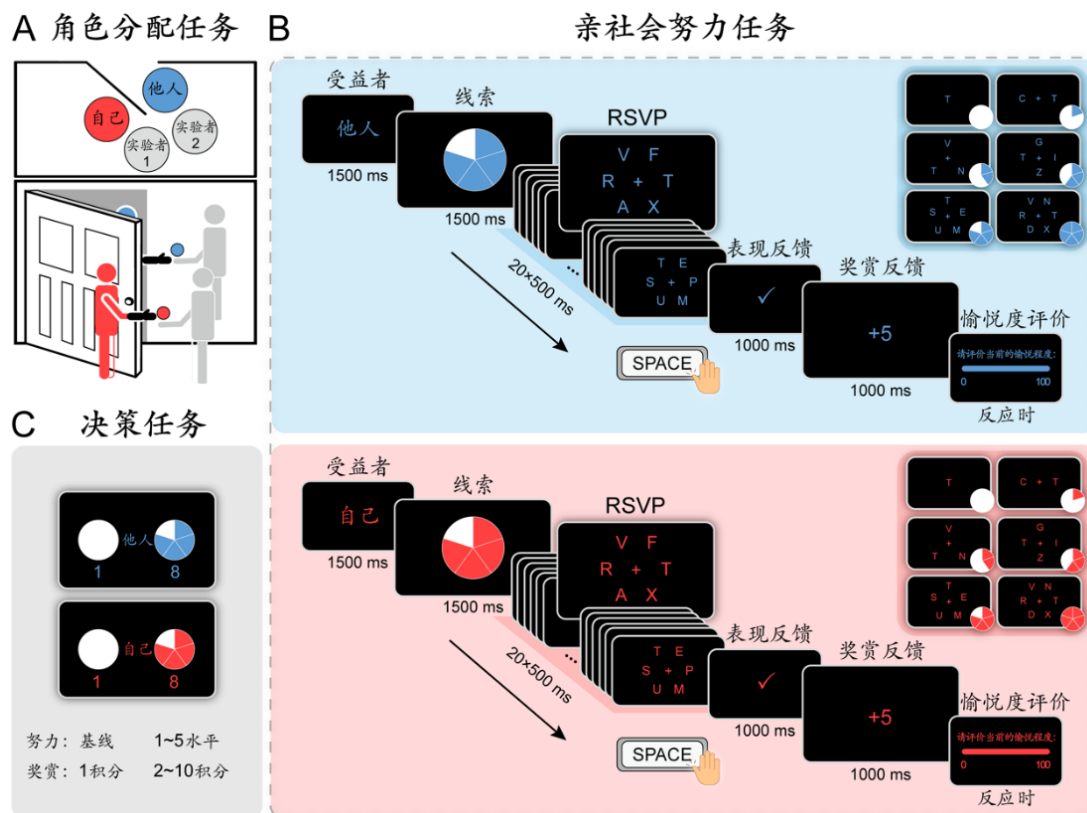


图 2 实验 1 任务流程图。

4.1.2 实验 2 亲社会努力影响奖赏加工的神经机制：EEG 和 fMRI 研究

在实验 1 的基础上,实验 2 旨在进一步揭示亲社会努力影响奖赏加工的神经动态过程和神经基础。该实验将采用与实验 1 类似的亲社会努力任务,并分别结合 EEG (实验 2a) 和 fMRI (实验 2b) 技术,对比被试在为自己和匿名陌生他人付出努力赢得奖赏时,在努力执行之前的奖赏线索评估和努力执行之后的奖赏反馈评估诱发的大脑反应。EEG 技术具有毫秒级的时间分辨率,能够追踪亲社会努力影响奖赏加工的完整神经动态变化过程,因此本研究将以 EEG 技术为主。另外,由于其在探测皮层下奖赏相关区域(如腹侧纹状体等)的局限性,有必要辅以 fMRI 技术(实验 2b)来考察努力执行前后相关大脑区域的神经反应,以获得完整的数据支撑。

实验 2 中我们将重点考察线索阶段与反馈阶段的神经反应差异,并结合多变量分析方法进一步比较利己努力与利他努力在神经表征上的异同。任务与实验 1 类似,并在以下方面进行修改:(1)被试成功完成努力任务以 50%概率获得 10 个奖赏积分、50%概率获得 0 个奖赏积分。我们采用概率性(而非确定性)奖赏以更好地诱发与奖赏相关的神经反应。若未能完成任务,则无法赢得奖赏。(2)我们调整试次内的时间信息和试次数以适用 EEG 和 fMRI 的实验设计。我们预期亲社会努力对奖赏加工的影响同样会在神经层面表现出双重贬值模式。在努力执行之前的线索阶段,线索 P3 和 CNV 反应以及前扣带回和前脑岛的神经激活随着努力水平的增加而降低,该折扣效应在为他人努力时更为明显;在努力执行之后的奖赏反馈阶段,RewP、反馈 P3 和 Delta 反应、以及腹侧纹状体和腹内侧前额叶皮层的神经激活,在为自己努力时随努力水平的增加而增加,在为他人努力时随努力水平的增加而降低。

4.2 研究二:亲社会努力影响奖赏加工的领域一般性/特异性机制

研究二通过 3 个 EEG 实验,从努力类型(实验 3)和收益领域(实验 4 和 5)两个维度,对比为自己和匿名陌生他人获益时努力执行前后的神经动态过程,系统考察亲社会努力影响奖赏加工的神经动态过程的领域一般性与特异性机制。在实验 3 中,我们采用对照实验设计,考察体力努力(肌肉收缩)和认知努力(注意资源)对亲社会奖赏加工的差异化影响。鉴于收益领域的多维特性,我们进一步设计了实验 4 和实验 5 以探讨不同收益领域亲社会努力影响奖赏加工的边界作用范围。实验 4 侧重于经济博弈奖惩领域的收益对比,比较付出努力为他人赢得奖赏和避免损失的神经动态过程;实验 5 重点探讨经济和道德领域的收益对比,比较付出努力为他人赢得奖赏和避免伤害的神经动态过程。

4.2.1 实验 3 领域一般性/特异性机制:体力和认知努力的对比研究

实验 3 旨在比较基于体力努力和认知努力的亲社会奖赏加工过程。研究将分别设置认知努力任务和体力努力任务,并在任务难度、时间成本等方面尽可能进行匹配,以比较两类努

力在为自己和为他人获有 50% 概率避免输掉奖赏积分（即损失 0 个积分），对努力执行前后的奖赏加工是否表现出一致的神经动态模式。认知努力任务与实验 2a 相似，体力努力任务以实验 2a 为基础进行修改。具体而言，我们将实验 2a 中的 RSVP 任务替换为握力计任务，要求被试完成不同体力努力水平任务。该实验有助于回答亲社会努力影响奖赏加工的机制究竟是跨努力类型普遍存在，还是在不同努力形式下具有特定表现。通过比较两类任务中的线索评估与反馈评估过程，本研究将进一步检验这一机制的领域一般性与特异性。我们预期亲社会努力对奖赏加工的影响在体力和认知努力之间表现出相似的双重贬值模式。在努力执行之前的线索阶段，利他努力比利己努力表现出更强的折扣效应，在努力执行之后的奖赏评估阶段，利己努力表现出增值效应，而利他努力表现出贬值效应。

4.2.2 实验 4 领域一般性/特异性机制：获得奖赏和避免损失的对比研究

实验 4 旨在比较获得奖赏和避免损失两类收益情境下，亲社会努力影响奖赏加工的神经动态过程，进一步检验不同收益领域亲社会努力影响奖赏加工的边界作用范围。研究将设置奖赏情境和损失情境，考察被试在为自己和为他人付出努力时，面对赢得收益与避免损失两类结果时的加工差异。在奖赏情境中，被试成功完成努力任务有 50% 概率获得 10 个奖赏积分，有 50% 概率获得 0 个奖赏积分；在损失情境中，被试成功完成努力任务有 50% 概率避免输掉奖赏积分（即损失 0 个积分），有 50% 概率输掉 10 个奖赏积分。在损失情境之前给予被试一定数量的奖赏积分作为为自己和为他人努力的本金。两种情境均要求被试分别为自己和他人完成努力任务，并在任务过程中呈现相应的线索与反馈信息。为保证两类情境具有可比性，本研究将在任务结构、努力水平、结果概率及反馈形式等方面尽量保持一致。我们预期亲社会努力在奖赏情境表现出双重贬值模式，在损失情境表现出悖论模式。与为他人赢得奖赏相比，为他人避免损失的亲社会努力机制向利己努力方向靠拢，表现为线索阶段的努力折扣效应减小，评估阶段出现努力增值效应。

4.2.3 实验 5 领域一般性/特异性机制：获得奖赏和避免伤害的对比研究

实验 5 旨在比较获得奖赏和避免伤害情境下，亲社会努力影响奖赏加工的神经动态过程。与实验 4 类似，研究 5 将在奖赏情境之外引入伤害回避情境，被试每次付出不同程度的努力为自己或匿名他人赢得回避疼痛刺激的机会。任务反馈将明确呈现本次是否成功避免伤害；对于未能避免的情况，将在部分试次中施加相应的疼痛刺激。为控制个体间痛觉敏感性的差异，研究将在正式实验前进行疼痛阈限测定，并据此对刺激强度进行个体化设定，以确保所有被试经历相似的痛觉感受。在避免伤害这一具有更强社会和道德意义的情境下，个体为他人付出努力时是否仍然表现出与奖赏情境相同的双重贬值模式，还是会因结果意义的变化而

出现不同的神经加工特点。通过比较线索阶段和反馈阶段的相关脑电指标,实验 5 将检验亲社会努力机制在不同收益领域中的一般性与特异性。我们预期亲社会努力在伤害情境表现出与利己努力相似、甚至是更为明显的努力悖论模式。

5 理论建构与创新

目前,亲社会行为的理论和实证研究大多聚焦于他人收益,较少关注努力成本的作用。现有少量关注努力成本的研究主要集中在体力努力和金钱收益的单一领域,仅考察了努力执行之前的价值决策过程,忽视了努力执行之后的价值评估过程。因此,本研究拟解决的核心科学问题是亲社会努力影响奖赏加工的认知神经反应在努力执行前后如何动态变化?为了回答这一问题,本研究创新性提出亲社会努力影响奖赏加工的双阶段模型,以期通过自我努力成本与他人收益的动态整合视角,系统揭示亲社会努力影响奖赏加工的认知神经机制,为促进亲社会行为提供新思路。

在研究视角上,本研究突破了传统亲社会行为研究对他人收益的单一关注,首次将努力成本及其动态加工过程纳入亲社会行为的研究框架,将亲社会行为界定为一个包含努力执行之前的价值决策和努力执行之后的价值评估的连续心理过程,从而超越了传统成本-收益框架的静态视角。同时,本研究基于前期利己努力的研究成果,创造性提出亲社会努力影响奖赏加工的双阶段模型,并进一步区分两种竞争性假说:一是双重努力贬值假说,二是先贬值后增值的努力悖论假说。根据该模型,当受益对象为自己时,利己努力表现出先贬值后增值的努力悖论模式;当受益对象为匿名陌生他人时,此时努力主要以预期成本的形式被表征,个体需要在努力负担与收益之间进行权衡,因此努力更可能降低结果的主观价值表现出折扣效应,形成双重努力贬值模式。

在研究设计上,本研究围绕上述理论框架构建了层层递进的研究方案。研究一结合行为、EEG 和 fMRI 等技术手段,分别从行为表现、神经动态过程和神经基础三个层面探究亲社会努力后效的基本机制,系统考察个体在为自己与为他人获益时,努力执行之前与努力执行之后奖赏加工的差异。该研究旨在回答两个关键问题:其一,自我努力成本如何影响他人收益的神经表征及其动态过程;其二,双阶段模型的双重努力贬值和努力悖论假说能否得到实证支持。在此基础上,研究二进一步从努力类型和收益领域两个维度考察上述机制的领域一般性与领域特异性,明确哪些效应具有跨领域稳定性,哪些受到努力类型和结果社会属性的调节,从而界定亲社会努力影响奖赏加工的适用范围。通过从基本机制到边界条件的递进式研究设计,本研究不仅能揭示亲社会努力的认知与神经机制,还能厘清其形成条件与作用边界,

进而为行为预测和干预提供依据。

总之，现实生活中许多人并非缺乏帮助他人的意愿，而是在需要持续付出努力时，更容易高估努力成本、低估结果价值，从而表现出亲社会淡漠。对这一现象，仅从社会规范角度解释往往难以触及心理根源。本研究从努力与奖赏加工的动态关系出发，厘清亲社会行为中自我努力成本对他人收益的影响，并验证双阶段模型的双重贬值和努力悖论假说，不仅有助于系统阐明亲社会努力影响奖赏加工的认知神经机制，也将为亲社会训练与干预实践提供理论依据。

参考文献

- 谢晓非, 寇贱. (2024). 亲社会行为: 社会互动和社会发展中的人性光辉(序言). *心理学报*, 56(9), 1157-1158. <https://doi.org/10.3724/sp.J.1041.2024.01157>
- Aronson, E., & Mills, J. (1959). The effect of severity of initiation on liking for a group. *The Journal of Abnormal and Social Psychology*, 59(2), 177-181. <https://doi.org/10.1037/h0047195>
- Batson, C. D., Batson, J. G., Slingsby, J. K., Harrell, K. L., Peekna, H. M., & Todd, R. M. (1991). Empathic joy and the empathy-altruism hypothesis. *Journal of Personality and Social Psychology*, 61(3), 413-426. <https://doi.org/10.1037/0022-3514.61.3.413>
- Bogdanov, M., Renault, H., LoParco, S., Weinberg, A., & Otto, A. R. (2022). Cognitive effort exertion enhances electrophysiological responses to rewarding outcomes. *Cerebral Cortex*, 32(19), 4255-4270. <https://doi.org/10.1093/cercor/bhab480>
- Brown, S. L., & Brown, R. M. (2015). Connecting prosocial behavior to improved physical health: Contributions from the neurobiology of parenting. *Neuroscience & Biobehavioral Reviews*, 55, 1-17. <https://doi.org/10.1016/j.neubiorev.2015.04.004>
- Bustamante, L. A., Oshinowo, T., Lee, J. R., Tong, E., Burton, A. R., Shenhav, A., Cohen, J. D., & Daw, N. D. (2023). Effort Foraging Task reveals positive correlation between individual differences in the cost of cognitive and physical effort in humans. *Proceedings of the National Academy of Sciences*, 120(50), e2221510120. <https://doi.org/10.1073/pnas.2221510120>
- Chong, T. T., Apps, M., Giehl, K., Sillence, A., Grima, L. L., & Husain, M. (2017). Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLoS Biology*, 15(2), e1002598. <https://doi.org/10.1371/journal.pbio.1002598>
- Clay, G., Mlynski, C., Korb, F. M., Goschke, T., & Job, V. (2022). Rewarding cognitive effort increases the intrinsic value of mental labor. *Proceedings of the National Academy of Sciences*, 119(5), e2111785119. <https://doi.org/10.1073/pnas.2111785119>
- Contreras-Huerta, L. S., Lockwood, P. L., Bird, G., Apps, M. A. J., & Crockett, M. J. (2022). Prosocial behavior is associated with transdiagnostic markers of affective sensitivity in multiple domains. *Emotion*, 22(5), 820-835. <https://doi.org/10.1037/emo0000813>
- Contreras-Huerta, L. S., Pisauro, M. A., & Apps, M. A. J. (2020). Effort shapes social cognition and behaviour: A neuro-cognitive framework. *Neuroscience & Biobehavioral Reviews*, 118, 426-439. <https://doi.org/10.1016/j.neubiorev.2020.08.003>
- Crockett, M. J., Kurth-Nelson, Z., Siegel, J. Z., Dayan, P., & Dolan, R. J. (2014). Harm to others outweighs harm to self in moral decision making. *Proceedings of the National Academy of Sciences*, 111(48), 17320-17325. <https://doi.org/10.1073/pnas.1408988111>
- Crockett, M. J., Siegel, J. Z., Kurth-Nelson, Z., Dayan, P., & Dolan, R. J. (2017). Moral transgressions corrupt neural representations of value. *Nature Neuroscience*, 20(6), 879-885. <https://doi.org/10.1038/nn.4557>
- Depow, G. J., Lin, H., & Inzlicht, M. (2022). Cognitive effort for self, strangers, and charities. *Scientific Reports*, 12(1), 15009. <https://doi.org/10.1038/s41598-022-19163-y>
- Dobryakova, E., Jessup, R. K., & Tricomi, E. (2017). Modulation of ventral striatal activity by cognitive effort. *NeuroImage*, 147, 330-338. <https://doi.org/10.1016/j.neuroimage.2016.12.029>
- Engel, C. (2011). Dictator games: A meta study. *Experimental Economics*, 14(4), 583-610. <https://doi.org/10.1007/s10683-011-9283-7>
- Forbes, P. A. G., Aydogan, G., Braunstein, J., Todorova, B., Wagner, I. C., Lockwood, P. L., Apps, M. A.

- J., Ruff, C. C., & Lamm, C. (2024). Acute stress reduces effortful prosocial behaviour. *eLife*, 12, RP87271. <https://doi.org/10.7554/eLife.87271>
- Gaule, A., Martin, P., Lockwood, P. L., Cutler, J., Apps, M., Roberts, R., Phillips, H., Brown, K., McCrory, E. J., & Viding, E. (2024). Reduced prosocial motivation and effort in adolescents with conduct problems and callous-unemotional traits. *Journal of Child Psychology and Psychiatry*, 65(8), 1061–1071. <https://doi.org/10.1111/jcpp.13945>
- Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 161–178.
- Hein, G., Silani, G., Preuschhoff, K., Batson, C. D., & Singer, T. (2010). Neural responses to ingroup and outgroup members' suffering predict individual differences in costly helping. *Neuron*, 68(1), 149–160. <https://doi.org/10.1016/j.neuron.2010.09.003>
- Hernandez Lallement, J., Kuss, K., Trautner, P., Weber, B., Falk, A., & Fließbach, K. (2014). Effort increases sensitivity to reward and loss magnitude in the human brain. *Social Cognitive and Affective Neuroscience*, 9(3), 342–349. <https://doi.org/10.1093/scan/nss147>
- Husain, M., & Roiser, J. P. (2018). Neuroscience of apathy and anhedonia: A transdiagnostic approach. *Nature Reviews Neuroscience*, 19(8), 470–484. <https://doi.org/10.1038/s41583-018-0029-9>
- Imas, A. (2014). Working for the "warm glow": On the benefits and limits of prosocial incentives. *Journal of Public Economics*, 114, 14–18. <https://doi.org/10.1016/j.jpubeco.2013.11.006>
- Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: Effort is both costly and valued. *Trends in Cognitive Sciences*, 22(4), 337–349. <https://doi.org/10.1016/j.tics.2018.01.007>
- Jiang, H., & Zheng, Y. (2024). Effort expenditure increases risk-taking for improbable rewards. *Personality and Social Psychology Bulletin*, 51(8), 1–15. <https://doi.org/10.1177/01461672231218746>
- Jones, B. A. (2021). A review of social discounting: The impact of social distance on altruism. *The Psychological Record*, 72(3), 511–515. <https://doi.org/10.1007/s40732-021-00488-5>
- Kool, W., & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour*, 2(12), 899–908. <https://doi.org/10.1038/s41562-018-0401-9>
- Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, 36(6), 661–679. <https://doi.org/10.1017/S0140525X12003196>
- Liu, J., Gu, R., Liao, C., Lu, J., Fang, Y., Xu, P., Luo, Y. J., & Cui, F. (2020). The neural mechanism of the social framing effect: Evidence from fMRI and tDCS studies. *Journal of Neuroscience*, 40(18), 3646–3656. <https://doi.org/10.1523/jneurosci.1385-19.2020>
- Lockwood, P. L., Abdurahman, A., Gabay, A. S., Drew, D., Tamm, M., Husain, M., & Apps, M. A. J. (2021). Aging increases prosocial motivation for effort. *Psychological Science*, 32(5), 668–681. <https://doi.org/10.1177/0956797620975781>
- Lockwood, P. L., Apps, M. A., & Chang, S. W. (2020). Is there a 'social' brain? Implementations and algorithms. *Trends in Cognitive Sciences*, 24(10), 802–813. <https://doi.org/10.1016/j.tics.2020.06.011>
- Lockwood, P. L., Hamonet, M., Zhang, S. H., Ratnavel, A., Salmony, F. U., Husain, M., & Apps, M. A. J. (2017). Prosocial apathy for helping others when effort is required. *Nature Human Behaviour*, 1(7), 0131. <https://doi.org/10.1038/s41562-017-0131>
- Lockwood, P. L., van den Bos, W., & Dreher, J.-C. (2025). Moral learning and decision-making across the lifespan. *Annual Review of Psychology*, 76(1), 475–500.

<https://doi.org/10.1146/annurev-psych-021324-060611>

- Lockwood, P. L., Wittmann, M. K., Nili, H., Matsumoto-Ryan, M., Abdurahman, A., Cutler, J., Husain, M., & Apps, M. A. J. (2022). Distinct neural representations for prosocial and self-benefiting effort. *Current Biology*, 32(19), 4172–4185. <https://doi.org/10.1016/j.cub.2022.08.010>
- Lopez-Gamundi, P., Mas-Herrero, E., & Marco-Pallares, J. (2024). Disentangling effort from probability of success: Temporal dynamics of frontal midline theta in effort-based reward processing. *Cortex*, 176, 94–112. <https://doi.org/10.1016/j.cortex.2024.03.014>
- Norton, M. I., Mochon, D., & Ariely, D. (2012). The IKEA effect: When labor leads to love. *Journal of Consumer Psychology*, 22(3), 453–460. <https://doi.org/10.1016/j.jcps.2011.08.002>
- Pan, W., Lu, J., Wu, L., Kou, J., & Lei, Y. (2023). Expending effort may share neural responses with reward and evokes high subjective satisfaction. *Biological Psychology*, 177, 108480. <https://doi.org/10.1016/j.biopsycho.2022.108480>
- Persson, E., Tinghög, G., & Västfjäll, D. (2024). Intertemporal prosocial behavior: a review and research agenda. *Frontiers in Psychology*, 15, 1359447. <https://doi.org/10.3389/fpsyg.2024.1359447>
- Pfattheicher, S., Nielsen, Y. A., & Thielmann, I. (2022). Prosocial behavior and altruism: A review of concepts and definitions. *Current Opinion in Psychology*, 44, 124–129. <https://doi.org/10.1016/j.copsyc.2021.08.021>
- Qu, C., Hu, Y., Tang, Z., Derrington, E., & Dreher, J. C. (2020). Neurocomputational mechanisms underlying immoral decisions benefiting self or others. *Social Cognitive and Affective Neuroscience*, 15(2), 135–149. <https://doi.org/10.1093/scan/nsaa029>
- Ruff, C. C., & Fehr, E. (2014). The neurobiology of rewards and values in social decision making. *Nature Reviews Neuroscience*, 15(8), 549–562. <https://doi.org/10.1038/nrn3776>
- Sellitto, M., Neufang, S., Schweda, A., Weber, B., & Kalenscher, T. (2021). Arbitration between insula and temporoparietal junction subserves framing-induced boosts in generosity during social discounting. *NeuroImage*, 238, 118211. <https://doi.org/10.1016/j.neuroimage.2021.118211>
- Shenhav, A., Musslick, S., Lieder, F., Kool, W., Griffiths, T. L., Cohen, J. D., & Botvinick, M. M. (2017). Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40, 99–124. <https://doi.org/10.1146/annurev-neuro-072116-031526>
- Suzuki, S., Lawlor, V. M., Cooper, J. A., Arulpragasam, A. R., & Treadway, M. T. (2021). Distinct regions of the striatum underlying effort, movement initiation and effort discounting. *Nature Human Behaviour*, 5(3), 378–388. <https://doi.org/10.1038/s41562-020-00972-y>
- Talbot, J., Cutler, J., Tamm, M., Little, S. J., Harmer, C. J., Husain, M., Lockwood, P. L., & Apps, M. A. J. (2025). Dopamine boosts motivation for prosocial effort in Parkinson's disease. *The Journal of Neuroscience*, e1593242024. <https://doi.org/10.1523/jneurosci.1593-24.2024>
- Umemoto, A., Lin, H., & Holroyd, C. B. (2022). Electrophysiological measures of conflict and reward processing are associated with decisions to engage in physical effort. *Psychophysiology*, e14176. <https://doi.org/10.1111/psyp.14176>
- Vanags, P., Cutler, J., Kosse, F., & Lockwood, P. L. (2025). Greater income and financial well-being are associated with higher prosocial preferences and behaviors across 76 countries. *PNAS nexus*, 4(2), pgae582. <https://doi.org/10.1093/pnasnexus/pgae582>
- Westbrook, A., van den Bosch, R., Maatta, J. I., Hofmans, L., Papadopetraki, D., Cools, R., & Frank, M. J. (2020). Dopamine promotes cognitive effort by biasing the benefits versus costs of cognitive work. *Science*, 367(6484), 1362–1366. <https://doi.org/10.1126/science.aaz5891>

- Wu, M., & Zheng, Y. (2023). Physical effort paradox during reward evaluation and links to perceived control. *Cerebral Cortex*, 33(15), 9343–9353. <https://doi.org/10.1093/cercor/bhad207>
- Wu, R., Ferguson, A. M., & Inzlicht, M. (2023). Do humans prefer cognitive effort over doing nothing? *Journal of Experimental Psychology: General*, 152(4), 1069–1079. <https://doi.org/10.1037/xge0001320>
- Yi, W., Mei, S., Zhang, M., & Zheng, Y. (2020). Decomposing the effort paradox in reward processing: Time matters. *Neuropsychologia*, 137, 107311. <https://doi.org/10.1016/j.neuropsychologia.2019.107311>
- Zentall, T. R. (2010). Justification of effort by humans and pigeons. *Current Directions in Psychological Science*, 19(5), 296–300. <https://doi.org/10.1177/0963721410383381>
- Zhang, M., & Zheng, Y. (2022). Neural dynamics of effort-modulated reward processing. *Psychophysiology*, 59(10), e14070. <https://doi.org/10.1111/psyp.14070>
- Zheng, Y., Zhang, M., & Wu, M. (2024). Effort discounts reward-based control allocation: A neurodynamic perspective. *Psychophysiology*, 61(2), e14451. <https://doi.org/10.1111/psyp.14451>

A two-stage model of the effects of prosocial effort on reward processing

ZHENG Ya, YANG Ziyang

(Center for Reward and Social Cognition, School of Education, Guangzhou University, Guangzhou 510006, China)

Abstract: Prosocial behavior often requires people to incur personal costs to benefit others. Although many studies have focused on the benefits received by others, the effects of self-incurred costs on prosocial behavior remain underexplored. The few studies that have examined self-costs have mainly focused on the trade-off between effort costs and monetary rewards, overlooking the broader neural dynamics through which effort shapes the valuation of benefits to others. Recently, we demonstrated an effort paradox mechanism, whereby effort exerted for oneself both prospectively discounts and retrospectively enhances reward value. We hypothesize that this two-stage relationship between effort and reward may constitute a novel mechanism underlying prosocial behavior. To test this hypothesis, we will combine behavioral experiments with EEG and fMRI to examine the neural dynamics that arise when individuals exert cognitive effort to earn rewards for themselves versus for others. We will further determine whether these neural dynamics are domain-general or domain-specific across different forms of effort and benefit. These findings will clarify the cognitive and neural mechanisms by which prosocial effort shapes reward processing, thereby addressing a key gap in current theories and offering a new perspective

on how self-incurred costs and others' benefits are balanced, and on how prosocial behavior can be promoted.

Keywords: prosocial behavior; cognitive effort; reward; effort paradox

ChinaXiv:202606.00002v1