道德判断影响大脑对公平行为的早期评价

占斌[†], 杜碧煊[†], 陈绍华, 李轶文, 何蔚祺*, 罗文波*

辽宁师范大学脑与认知神经科学研究中心, 大连 116029

[†] 同等贡献

* 联系人, E-mail: hewq@lnnu.edu.cn; luowb@lnnu.edu.cn

2019-12-15 收稿, 2020-01-10 修回, 2020-01-10 接受, 2020-01-13 网络版发表

国家自然科学基金(31970991, 31871106)资助

摘要 近年研究发现, 社会距离、社会排斥、社会评价等社会因素与公平行为共同调节了早期结果评价过程, 但目前研究尚未探讨社会因素中道德判断对公平决策的影响及其神经机制. 本研究在道德句子启动下与最后通牒范式结合并记录脑电数据, 探究道德判断(3水平: 杀人犯、献血者、普通人)与提议公平性(2水平: 公平、不公平)对响应者的决策行为影响. 行为结果发现, 献血者不公平提议的接受率显著高于杀人犯, 杀人犯公平提议的接受率显著低于献血者与普通人. 脑电结果发现, 道德判断和提议公平性共同调节了个体的FRN(feedback-related negativity)波幅, 仅发现献血者条件下不公平提议比公平提议诱发更负的FRN波幅. 在P300成分上发现公平比不公平提议诱发了更大的波幅, 且不受道德判断的影响. 本研究结果表明, 道德判断与公平准则冲突时, 个体更依赖于道德判断进行决策. 个体对高道德水平者存在更高的公平期待, 对其不公平行为诱发了更强的负性情绪. FRN反映了违反社会准则的不公平行为, 这一结果早期评价过程也受到了道德判断等社会因素的调节作用, 但P300反映的晚期评价过程则不受道德判断的影响.

关键词 道德判断, 最后通牒, 反馈相关负波(FRN), P300, 公平准则

道德作为社会文化中的重要内容, 是个体如何对待他人的重要准则, 包括如何处理个体与他人之间的正义、公平和权利等^[1,2]. 道德判断(moral judgment)是在社会准则下对自我和他人行为进行评价的过程, 既包含一般认知过程, 也包括情绪过程^[1-4]. 道德判断涉及个体帮助或伤害他人的行为, 影响个体与他人的社会互动. 例如, Cui等人^[5]研究发现, 道德判断影响了对他人的疼痛知觉. 个体观看高道德水平者的疼痛图片比中性图片诱发更大的N2波幅, 而观看低道德水平者的疼痛图片诱发的N2波幅与中性图片无差异, 表现出个体对不道德者疼痛共情的减弱. 同时研究发现, 个体对于道德评价更高的他人, 会有更高的道德期待, 当高道德个体做出不道德行为时, 人们对不道德行为的厌

恶评分更高^[6]. 研究者认为, 道德与社会合作的潜在通路是个体对高道德者存在信任偏向, 信任促进社会合作^[7,8]. 上述研究表明, 道德判断可能会影响个体对他人的共情、社会期待与信任等社会认知过程.

不患寡而患不均, 公平感是影响对他人社会认知中另一重要因素. 在传统经济学理论中, 个体被认为是理性的, 追求个人利益的最大化. 然而大量研究发现, 个体在进行决策时, 往往受到不公平感等主观因素的影响^[9,10]. 研究者利用最后通牒范式(ultimatum game, UG)研究个体的公平决策行为^[11,12]. UG任务是由两名参与者分配一定数额的金钱, 其中一人为“提议者”, 提议金钱的分配方案; 另一人为“响应者”, 选择接受或拒绝提议者提出的分配方案. 接受提议, 则按照该提议进

引用格式: 占斌, 杜碧煊, 陈绍华, 等. 道德判断影响大脑对公平行为的早期评价. 科学通报, 2020, 65: 1985–1995

Zhan B, Du B X, Chen S H, et al. Moral judgment modulates fairness consideration in the early outcome evaluation stage (in Chinese). Chin Sci Bull, 2020, 65: 1985–1995, doi: 10.1360/TB-2019-0681

行金额的分配, 拒绝提议, 则两人均可得不到金钱。结果发现, 个体并不是按照利益最大化的理性原则提出分配方案, 提议者往往会将一半左右的金额分配给对方, 而响应者通常会拒绝少于总金额中20%的提议^[9,13]。进一步研究发现, 响应者的公平决策受到社会情境的影响, 如个体明确知道自己在与电脑进行博弈时, 对不公平提议的拒绝率显著低于与真实人类进行的博弈^[11,14,15]。因此, 研究者认为, 在UG中响应者拒绝不公平提议反映了个体在社会决策情境下的公平偏好。

使用高时间分辨率的事件相关电位技术(event-related potential, ERP)研究UG任务中大脑对不同提议类型的时间进程, 发现反馈相关负波(feedback-related negativity, FRN)和P300成分反映了个体对提议公平性的不同评价阶段^[14,16,17]。FRN是指在反馈刺激呈现后200~350 ms的时间窗出现的一个负走向的脑电成分。溯源定位发现, FRN主要发生在前扣带皮层(anterior cingulate cortex, ACC), 与认知控制、行为决策和冲突监控相关^[18-21]。早期强化学习理论认为, 反馈学习过程中FRN表征了预期奖赏与实际奖赏之间的误差。当实际奖赏大于预期奖赏时, 为正性反馈, 反之为负性反馈。负性反馈会诱发更大的FRN波幅, 反映了反馈加工中认知学习过程^[18,22]。溯源分析结果也发现, FRN与情绪加工的ACC前部有关, FRN也反映了结果所引发的情绪动机差异^[23]。近年来, 大量的研究逐渐认识到二者是不可分离的, FRN既离不开对于奖赏预期的加工, 也无法摆脱情绪动机对认知过程的影响^[24-26]。

UG任务中, 不公平提议比公平提议诱发更负的FRN。FRN波幅也受到社会距离、面孔吸引力、社会排斥和社会评价等社会因素的影响^[27-29]。朋友提出的不公平提议比公平提议诱发更大的FRN, 而陌生人的提议公平与否并未引发FRN波幅变化, 表明社会距离调节FRN在不同公平条件下的波幅^[17]。男性作为响应者, 提议者为高吸引力女性面孔时, 不同提议类型诱发的FRN波幅无差异, 而在低吸引力面孔下, 不公平提议比公平提议诱发更负的FRN^[30]。在使用掷球游戏创造的情境中发现, 社会排斥影响个体公平决策的早期结果评价过程, 排斥者提出的不公平方案会诱发更大的FRN^[31]。FRN波幅还受到社会评价的影响。研究者将提议类型和面孔进行匹配, 总是提出不公平分配的面孔相较总是提出公平分配的面孔诱发更大的FRN, 不公平提议给面孔赋予了负性价值^[32]。FRN表征了强化学习过程中结果的好坏, 也反映了不同社会因素对早期

结果评价过程的影响。

反馈刺激出现后300~600 ms, 在顶枕区出现正走向的P300成分。P300成分与选择性注意、资源分配等高级认知过程相关^[33-35]。在UG任务中, P300反映了晚期加工中对结果的评估过程, 不仅表征结果的大小, 而且对奖赏幅度敏感^[36,37]。Wu和Zhou^[38]通过对奖赏效应、奖赏幅度、奖赏预期的控制, 发现奖赏的金额越大, P300所诱发的波幅越大, 并且符合预期的正性奖赏能够诱发更大的P300。因此, 在UG任务中公平提议比不公平提议诱发了更大的P300波幅, 公平提议符合社会合作下的公平准则且相比不公平提议获得的奖赏更大^[38-40]。

大量研究发现, 多种社会因素影响公平决策的早期评价阶段, 道德判断这一重要的社会因素也影响着人们的决策行为^[6,41]。Delgado等人^[7]发现, 在信任游戏中, 先验的道德信息可以调节经济决策中的合作行为。人们更容易相信先前具有高道德信息的陌生人, 在结果评价阶段对高道德者的信任偏向调节了尾状核的激活。Hu等人^[42]在信任游戏中发现, 相较于不信任决策, 信任决策的负性反馈比正性反馈诱发更大的FRN波幅, 个体选择信任决策, 则有更强烈的动机和期待获得正性奖赏。Boksem和De Cremer^[27]进一步发现, 个体的公平感更强时, 不公平提议诱发的FRN波幅更负。个体是否会通过伤害他人获取个人利益与关心他人的福利是道德决策的核心。研究发现, 大多数人避免伤害他人来获取金钱利益, 且愿意牺牲更多的金钱减轻陌生人的痛苦, 表现出牺牲个人利益以避免他人受苦的道德行为^[43,44]。人们基于自身道德信念牺牲物质利益来支持或者反对某些社会行为, 如人们的捐赠行为或拒绝欺骗他人获取利益等行为。当道德选择战胜了自私物质利益时, 前额叶皮层前侧显著激活^[45]。上述研究表明, 道德判断与公平偏好分别影响人们的决策行为, 但道德判断是否会影响人们的公平偏好, 进而影响人们的决策行为, 及其相应的神经动态变化机制尚不明确。ERP技术拥有高时间分辨率的特点, 适用于研究个体决策随着时间动态变化过程。本研究目的在于探讨道德判断是否与社会排斥、社会距离等社会因素类似也会影响公平决策的早期评价过程以及道德判断如何影响个体在经济决策中的公平偏好。

本实验在道德句子启动下进行UG任务, 来考察道德判断与公平偏好如何影响响应者的决策行为及其相应的神经动态变化过程。基于前人的研究结果, 本研究

进一步探讨个体对他人的道德判断(杀人犯、普通人、献血者)如何影响个体对于提议类型(公平、不公平)的接受程度。本研究假设个体更愿意接受献血者的不公平提议,而更多拒绝杀人犯的提议。在ERP结果上,假设不公平提议会比公平提议诱发更负的FRN,更重要的是,与其他社会因素相似,道德判断会调节与结果评价相关的FRN波幅。由于不同提议类型所产生的奖赏幅度不同,P300成分可能会受到提议类型的影响。

1 方法

(i) 被试。本实验招募25名(12男13女)在校本科生,年龄18~25周岁,平均年龄20.36岁($SD=1.82$)。所有被试均为右利手,视力正常或矫正视力正常,自我报告无精神或神经病史,近期末参与过类似实验。该实验得到校伦理委员会审核通过,实验开始前所有被试均签署了知情同意书,实验后给予一定报酬。使用GPower3.1先验计算被试量,效应量 $f=0.25$,统计检验力 $1-\beta=0.90$, $\alpha=0.05$ 前提下,进行 F 检验被试内重复测量方差分析,得到所需的被试量为24人。

(ii) 实验流程。实验在一个昏暗且安静的房间中进行,被试坐在舒适的椅子上,19 in(1 in=2.54 cm)、60 Hz刷新率的彩色屏幕放置于被试前。程序使用E-Prime 2.0软件编写,所有实验刺激材料均在屏幕中央位置呈现。本实验采用UG范式,3×2被试内设计。第一个因素为道德判断(3水平:杀人犯、献血者、普通人),参照Cui等人^[5]研究,该因素由道德句子启动,分别为

“他是一名杀人犯”、“他是一名献血者”、“他是一名普通人”。实验开始前,告知被试将与不同的人进行最后通牒博弈。为了帮助被试更好地融入实验情境,实验前给予被试两则新闻材料,杀人犯的材料是在火车站暴徒杀害无辜乘客,献血者的材料是一名普通大学生献血的事迹,普通人无道德信息材料。要求被试仔细阅读新闻材料后,分别对杀人犯、献血者、普通人进行道德感评分。对三类人的道德评分进行单因素的方差分析, $F(2,72)=319.492$, $P<0.001$, $\eta_p^2=0.899$ 。事后比较发现,被试对献血者($M\pm SD$, 7.720 ± 0.891)的道德评分显著高于普通人(5.880 ± 0.726 , $P<0.001$)与杀人犯(1.480 ± 1.046 , $P<0.001$),而对普通人的道德评分也显著高于杀人犯($P<0.001$)。第二个因素为提议类型(2水平:公平、不公平),本实验有5种分配方案,总金额为50元,分配给被试¥5,¥10,¥15,¥20和¥25。其中¥5和¥10为不公平提议,¥20和¥25为公平提议,¥15作为中等公平条件,不纳入脑电分析。实验流程见图1,首先呈现400~600 ms注视点,接着呈现1500 ms道德句子启动,以提示被试本轮游戏中提议者的道德类型,随后呈现分配金额,被试(响应者)有充足的时间进行反应。选择按键“F”,同意该提议,则按照该提议分配金额。按键“J”,拒绝该提议,则两人都得不到钱(¥0)。被试按键反应后金额分配界面消失,呈现1500 ms的结果反馈。刺激间隔(inter-stimulus interval)均为500 ms,试次间隔为200~300 ms。被试在实验中所得到的总金额按照一定比例的折算作为被试报酬。在3种水平的道德信息下,5

图1 实验流程示意图

Figure 1 Illustration of a representative experimental trial

种提议方案各重复30次,共450个试次,分为6个block.

(iii) EEG记录. 使用国际标准10-20系统扩展64导电极帽,以BP系统(Brain Products, Munich, Germany)记录脑电数据. 记录时,所有电极点的电阻均低于5 k Ω ,在线参考点为FCz,离线参考转为全脑平均参考. 同时,在右眼眶下1 cm处放置电极点采集垂直眼电(VEOG). 连续记录脑电时带通滤波为0.01~100 Hz,采样率为1000 Hz/导.

(iv) 脑电数据处理与统计分析. 本实验采用MATLAB 2018b与EEGLAB工具箱(在MATLAB环境下运行的开源工具包)^[46],分析离线数据. 首先,带通滤波为0.01~30 Hz,原始脑电数据以提议方案界面开始前200~1000 ms进行分段,并对分段数据进行基线校正. 接着,使用独立成分分析^[46](independent component analysis, ICA),移除与眼电相关的独立成分(1~2个),剔除超过 ± 80 μ V的试次. 将提议方案为¥5与¥10的试次叠加平均,计算为不公平提议,¥20与¥25元的试次叠加平均,计算为公平提议. 本研究根据地形图分布与前人研究^[17,30],选取了FRN和P300成分进行统计分析. 根据合并定位(collapsed localizer)策略^[47,48]对每个ERP成分选取了不同的电极点与时间窗计算平均波幅. FRN成分选取了F1, Fz, F2, FC1, FCz, FC2, C1, Cz和C2,时间窗为300~350 ms; P300选取了CP3, CP4, CPz, P3, P4与Pz,时间窗为450~600 ms. 所有ERP成分使用两因素重复测量方差分析进行统计分析,第一个因素为道德判断(3水平);第二个因素为提议类型(2水平). 行为结果将5种提议分配方案的接受率作为因变量,使用单因素方差分析,研究不同道德判断对不同分配的接受率. 所有统计分析均采用了Greenhouse-Geisser epsilon校正 P 值与自由度,事后比较均采用了Bonferroni校正($P < 0.05$),并使用 η_p^2 来估计效应量. 以上统计分析均使用SPSS 22.0软件进行分析.

2 结果

2.1 行为结果

以5种分配方案的接受率作为因变量,道德启动作为自变量,进行单因素方差分析. 在提议为¥5时, $F=4.741$, $P=0.012$, $\eta_p^2=0.116$,与¥10时, $F=6.788$, $P=0.002$, $\eta_p^2=0.159$,事后比较发现献血者的接受率显著高于杀人犯的接受率($P < 0.009$),普通人与献血者不存在差异($P > 0.221$),普通人与杀人犯不存在差异

($P > 0.197$). 提议为¥15, $F=9.153$, $P < 0.001$, $\eta_p^2=0.203$,提议为¥20, $F=7.794$, $P=0.001$, $\eta_p^2=0.178$,与提议为¥25时, $F=12.524$, $P < 0.001$, $\eta_p^2=0.258$,事后比较发现献血者($P < 0.001$)与普通人($P < 0.026$)均显著高于杀人犯的接受率,献血者与普通人不存在差异($P > 0.121$). 行为结果见图2.

2.2 ERPs结果

(i) FRN结果. FRN统计结果显示,提议类型的主效应显著, $F(1,24)=18.816$, $P < 0.001$, $\eta_p^2=0.439$,道德判断主效应不显著, $F(2,48)=0.277$, $P=0.759$, $\eta_p^2=0.011$. 事后检验比较发现,不公平提议($M \pm SE$, -1.557 ± 0.480 μ V)比公平提议(-1.199 ± 0.442 μ V, $P < 0.001$)诱发更负的FRN波幅,见图3(a). 提议类型与道德判断存在显著的交互作用, $F(2,48)=3.553$, $P=0.037$, $\eta_p^2=0.129$,简单效应分析结果显示,仅在献血者的条件下, $F(1,24)=17.71$, $P < 0.001$, $\eta_p^2=0.425$,不公平提议(-1.784 ± 0.535 μ V)诱发的FRN波幅比公平提议(-1.066 ± 0.461 μ V, $P < 0.001$)更负,而在杀人犯与普通人的条件下,提议类型差异均不显著($P > 0.121$),见图4.

(ii) P300结果. P300结果显示,提议类型的主效应显著, $F(1,24)=7.168$, $P=0.013$, $\eta_p^2=0.230$,道德判断主效应不显著, $F(2,48)=0.466$, $P=0.630$, $\eta_p^2=0.019$,道德判断与提议类型不存在交互效应, $F(2,48)=0.304$, $P=0.304$, $\eta_p^2=0.013$. 事后检验发现公平提议($M \pm SE$, 2.124 ± 0.353 μ V)比不公平提议(1.759 ± 0.381 μ V, $P=0.013$)诱发更大的P300波幅,见图3(b).

3 讨论

本实验通过ERP技术与最后通牒范式研究在不同

图2 (网络版彩色)行为结果折线图
Figure 2 (Color online) Behavioral results

图3 (网络版彩色)FRN(a)与P300(b)提议类型主效应波形图与差异地形图

Figure 3 (Color online) Main effect ERPs and the topographical map of difference wave of FRN (a) and P300 (b)

图4 (网络版彩色)FRN交互作用波形图与差异地形图

Figure 4 (Color online) Interaction effect ERPs and the topographical map of difference wave of FRN

道德判断下, 响应者对公平与不公平提议的决策过程及其相应的神经动态机制。行为结果表明, 献血者不公平提议(¥5与¥10)的接受率显著高于杀人犯, 杀人犯的公平提议(¥20与¥25)的接受率远低于献血者与普通人。说明当道德判断与公平准则发生冲突时, 人们更倾向于依赖道德信息, 而不是公平与否来进行决策, 即能接受道德的人做出不公平分配, 面对不道德的人, 即使做出公平分配时, 也倾向于拒绝, 表现出惩罚不道德的人。ERPs结果显示, 相较于公平提议, 不公平提议诱

发更负的FRN, 该结果与前人的研究结果一致^[13,17,30]。同时, 结果评价相关的FRN波幅受到提议类型与道德判断共同调节, 具体来说, 仅在献血者条件下, 不公平比公平提议诱发了更负的FRN波幅, 在普通人与杀人犯条件下, 均未发现提议类型对FRN波幅的影响。在P300成分上发现, 相较于不公平提议, 公平提议诱发更大的P300, 而不受道德判断调节。本研究为早期结果评价阶段, 道德判断影响个体公平决策提供了新的证据。

本实验的行为结果证实了之前研究中的不公平厌

恶和利他性惩罚^[49],还发现了道德判断影响公平决策的证据.具体而言,相较于杀人犯,个体更愿意接受献血者的不公平提议,甚至拒绝杀人犯的公平提议.究其原因,可以通过道德与奖惩、社会情感的关系角度进行解释.首先,道德判断包含了某个人应该受到奖赏或者惩罚的信念^[50],并且道德判断激活了与奖赏、惩罚相关的腹内侧前额叶(ventromedial PFC, VMPFC)^[51,52].面对杀人犯时,个体会产生强烈的利他性惩罚和道德性惩罚,拒绝与其进行合作,而面对献血者时,则更多采取合作行为.经济学^[53]和进化心理学^[4,54]认为道德是经济合作的基础,道德判断通过增加信任和诚信来促进经济合作^[8].在遭遇冲突情境时,道德判断影响个体的决策行为,主要表现为帮助有道德的个体,惩罚没有道德个体^[55,56].本研究中不同道德判断所导致的公平行为差异,可能受到道德相关合作或惩罚的调节.其次,根据社会直觉模型^[57],情感是个体道德判断的重要驱动力,道德情感甚至比其他情感更加重要^[58].Cui等人^[5]发现,道德判断能够调节对他人的疼痛知觉,其内在机制为减弱对不道德个体的共情.不道德行为诱发个体厌恶等负性情绪^[59,60],道德行为使个体产生感恩等积极情绪.厌恶^[61,62]等负性情绪^[63]提高了个体拒绝不公平提议行为,而感恩等积极情绪能促进经济合作,即使是以牺牲个人更大的经济利益为代价^[64].当提议者为杀人犯时诱发个体的负性情绪,而献血者能诱发个体的积极情绪,通过情绪调节个体的公平决策.需要注意的是,本研究未发现普通人在不公平条件下(¥5和¥10)与杀人犯和献血者的差异,该结果可能是由于不公平提议本身存在较低的接受率,因此产生了地板效应.本研究结果揭示,当他人道德水平较高时,合作意向和积极情绪促进了个体的合作行为,更多的接受不公平行为.而当他人道德水平较低时,惩罚意向和负性情绪增加了个体的利他性惩罚.因此,当道德判断与公平准则冲突时,个体更依赖道德认知做出决策.

FRN表征了强化学习中的奖赏预测误差(reward prediction error, RPE),即预期结果与实际结果之间的差异,负性结果会比正性结果产生更大的FRN,意料之外的结果会比预期的结果产生更负的FRN.本研究与前人研究结果一致^[17,27],在FRN成分上发现,不公平提议比公平提议诱发更负的FRN波幅.公平准则是长期建立的社会规范,不公平提议违背了个体的公平偏好^[65,66],会使个体产生预期误差,从而诱发更负的FRN.本实验还发现了在献血者条件下,提议类型对FRN波

幅的调节作用,然而该效应并未出现在普通人与杀人犯条件下.FRN波幅会受到社会评价与情境的影响^[32],研究发现,社会距离与公平行为共同调节了FRN,与朋友进行博弈时,出现了不公平条件下的FRN大于公平条件,而与陌生人的博弈中并未发现这一差异.在社会情境下,增加了对朋友的公平期待,减弱了对陌生人的公平期待,因此在陌生人条件下提议类型在FRN上的调节效应消失^[17].根据前人研究中社会因素对FRN的调节作用^[30,31],本研究认为与社会距离调节FRN效应类似^[17],道德是社会准则和社会期望的重要组成部分^[67],人们对高道德水平者有更高的社会期待.本实验中道德句子启动了个体对不同道德水平提议者不同程度的公平期待.献血者在社会评价中具有高道德属性,个体对其产生强烈的公平期待,而减弱了普通人和杀人犯的公平期待.当献血者提出不公平分配时违反公平期待,造成了响应者的预期误差和认知冲突,从而诱发更负的FRN波幅.本实验结果与信任决策研究结果一致^[42],相比于不信任决策,信任决策下负性结果比正性结果诱发了更大的FRN,表明在信任决策下,个体更期待正性的结果反馈.除此之外,FRN的动机情绪理论认为,FRN反映负性反馈引发的情绪动机的评价过程^[21,23].道德能够增加与结果评价相关的注意与动机过程^[68].献血者提出不公平分配造成的负性情绪比普通人与杀人犯更强烈,导致了献血者条件下不公平提议诱发的FRN波幅更负.FRN的溯源结果在ACC,ACC与冲突监控功能相关^[21].研究者进一步发现,ACC的激活与违反社会期待的认知加工过程相关^[69].FRN的强化学习理论中也认为,FRN并不是由结果的绝对效价来决定,而是结果的效价与期望共同调节^[18,70].综上,本实验中道德判断与提议公平性的交互效应表明,在早期结果评价过程中个体对高道德水平者会产生更加强烈的动机与期望获得公平分配的结果.

有趣的是,献血者不公平比公平行为诱发了更大的FRN,该差异未出现在普通人与杀人犯中.而在行为结果上,献血者的不公平行为,相比于杀人犯更能被接受,甚至拒绝杀人犯的公平提议.即个体有更强烈的动机和期待高道德水平者的公平分配,这看似与较多接受高道德水平者的不公平行为,较多拒绝低道德水平者的公平提议的决策行为相冲突,实则体现了道德奖赏与经济奖赏的相互补偿.个体接受献血者的不公平提议,表现出牺牲个人利益为他人增加福利的道德行为^[43,71],而拒绝杀人犯的公平提议,也表现出牺牲经济

利益惩罚坏人的道德感^[50,72]。研究发现,对他人违反社会准则行为进行惩罚时,左侧DLPFC与眶额叶皮层(orbitofrontal cortex)显著激活,该脑区同样在进行道德判断任务下激活^[72]。个体对不同道德水平者提议的接受或拒绝行为本身具有道德奖赏。已有研究表明,在权衡道德与经济分配的任务中,人们愿意牺牲经济利益换取道德价值,主要激活VMPFC脑区^[7,71]。因此,接受献血者的不公平行为所引起的道德奖赏补偿了其金钱分配不公平的代价及其不公平分配诱发的负性情绪。反之,拒绝杀人犯的公平提议导致的经济损失被个体惩罚不道德者的行为引起的道德奖赏所弥补。经济学理论认为,人的亲社会行为由个体对金钱和与社会评价相关的利他主义(如道德价值)决定^[73],这与本研究的结果相符。本研究中道德判断不仅激活了个体对高道德者的公平期待^[32,42,69],且激活了个体的道德价值信念^[50]。FRN作为早期的结果自动评价过程,反映了个体对不同道德水平的公平行为预期,而个体的决策行为则由道德奖赏与金钱分配的公平性共同决定,且更依赖于道德价值进行决策。

在P300成分上,本研究发现公平提议比不公平提议诱发更大的波幅,而道德判断对P300无调节作用。一般来说,P300被认为反映了晚期结果评价中动机^[35]和注意分配过程^[74]。研究发现,与大脑早期对负性情绪敏感不同,晚期评价过程更倾向于加工积极情绪刺激,与负性刺激相比,积极刺激可以诱发更大的P300^[75,76]。奖赏加工与情绪加工类似,晚期阶段正性反馈比负性反馈获得更多注意加工。本研究结果与已有研究一致,提议者提出公平提议符合社会合作下的公平准则且相比不公平提议获得的金钱奖赏更大,而不公平提议则更多地与负性情绪与负性反馈相关^[30,38]。此外,研究者认为预期外结果会诱发P300成分的差异^[40]。P300波幅变化反映了意外事件的监控加工过程。基于这一观点,在公平及不公平方案出现概率相同的情况下,P300波幅在不同条件下应不存在差异。因此,本研究中发现的P300成分在不同提议类型间的差异,表明了经济决策任务中的公平偏好。在不同道德判断条件下,未发现P300成分的差异。这与前人在社会距离以及道德信息的UG任务中所得结果一致^[17,41]。前人研究发现,P300反映了高级认知加工中对自我相关刺激的选择性注意过程^[33]。本研究中提议类型会导致与个人利益紧密相

关的金钱奖赏差异,而道德判断属于对他人的评价过程,因此,P300晚期阶段更多参与自身金钱奖赏密切相关的选择性注意加工过程而不受到对他人道德评价过程的影响。此外,FRN反映的是对奖赏、公平与道德等多种因素综合的价值评价,而P300则是对一般的奖赏的效价敏感。道德价值与其他类型的价值均通过多巴胺环路进行表征^[45,77],研究者使用动态因果模型(dynamic causal modeling)证明了FRN的波幅直接受到来自中脑多巴胺神经元投射至纹状体与额叶的REP影响^[78],因此,道德价值与金钱奖赏通过多巴胺环路共同影响了FRN波幅,而P300不受到来自多巴胺神经环路的道德价值影响。综上所述,在晚期加工阶段中,大脑对提议的公平性与奖赏幅度敏感,而不受道德判断的影响。

本研究也存在一些不足,实验中所操纵道德判断的生态效度不高,这可能会影响被试在实验中的情境代入。本实验中仅研究了道德判断对劣势不公平条件的影响,未考虑对优势不公平条件的影响,尚未使用道德相关量表对被试的道德水平进一步控制。未来研究中,可以采取更高生态效度的道德情境进行诱发,进一步探讨道德判断对优势不公平决策的影响及个体化差异。此外,未来研究应关注人们在权衡道德价值与公平决策时的脑机制以及利用合理的计算模型量化道德价值、金钱奖赏、公平准则背后的神经计算机制。

4 结论

本实验通过道德判断与最后通牒范式相结合,探讨了道德判断对响应者公平决策的影响及其相关的脑电活动。行为结果发现,响应者更能接受高道德水平者的不公平提议,同时,更多拒绝低道德水平者的公平提议。道德判断与公平准则相冲突时,响应者更依赖于道德判断做出决策。ERP结果发现,仅在献血者条件下,不公平提议诱发的FRN波幅大于公平提议,表现出对献血者更高的公平期待,及其不公平行为诱发更为强烈的负性情绪。此外,公平提议的P300波幅比不公平提议波幅更大,但不受到道德判断的调节。本实验结果表明,违反公平准则的不公平行为能被早期的FRN成分所快速探测,并且这一早期快速的评价过程受到了道德判断等社会因素的调节,但P300反映的晚期结果评价过程则不受道德判断的调节。

参考文献

- 1 Yoder K J, Decety J. The good, the bad, and the just: Justice sensitivity predicts neural response during moral evaluation of actions performed by others. *J Neurosci*, 2014, 34: 4161–4166
- 2 Wang Y Q, Guo B Y. Is morality hardwired into the brain? The recent advances and prospects of cognitive neural mechanism researches of morality (in Chinese). *Chin Sci Bull*, 2017, 62: 2867–2875 [王云强, 郭本禹. 大脑是如何建立道德观念的: 道德的认知神经机制研究进展与展望. 科学通报, 2017, 62: 2867–2875]
- 3 Greene J D, Sommerville R B, Nystrom L E, et al. An fMRI investigation of emotional engagement in moral judgment. *Science*, 2001, 293: 2105–2108
- 4 Ridley M. *The Origins of Virtue*. Cardiff: Penguin Books Limited, 1997
- 5 Cui F, Ma N, Luo Y J. Moral judgment modulates neural responses to the perception of other's pain: An ERP study. *Sci Rep*, 2016, 6: 20851
- 6 Lu J, Peng X, Liao C, et al. The stereotype of professional roles influences neural responses to moral transgressions: ERP evidence. *Biol Psychol*, 2019, 145: 55–61
- 7 Delgado M R, Frank R H, Phelps E A. Perceptions of moral character modulate the neural systems of reward during the trust game. *Nat Neurosci*, 2005, 8: 1611–1618
- 8 Simpson B, Harrell A, Willer R. Hidden paths from morality to cooperation: Moral judgments promote trust and trustworthiness. *Social Forces*, 2013, 91: 1529–1548
- 9 Sanfey A G, Rilling J K, Aronson J A, et al. The neural basis of economic decision-making in the ultimatum game. *Science*, 2003, 300: 1755–1758
- 10 Rabin M. Incorporating fairness into game-theory and economics. *Am Econ Rev*, 1993, 83: 1281–1302
- 11 van't Wout M, Kahn R S, Sanfey A G, et al. Affective state and decision-making in the ultimatum game. *Exp Brain Res*, 2006, 169: 564–568
- 12 Güth W, Schmittberger R, Schwarze B. An experimental analysis of ultimatum bargaining. *J Econ Behav Organ*, 1982, 3: 367–388
- 13 Wu Y, Zhou Y, van Dijk E, et al. Social comparison affects brain responses to fairness in asset division: An ERP study with the ultimatum game. *Front Hum Neurosci*, 2011, 5: 131
- 14 Mayer S V, Rauss K, Pourtois G, et al. Behavioral and electrophysiological responses to fairness norm violations in antisocial offenders. *Eur Arch Psychiatry Clin Neurosci*, 2019, 269: 731–740
- 15 Chen M, Zhao Z, Lai H. The time course of neural responses to social versus non-social unfairness in the ultimatum game. *Soc Neurosci*, 2019, 14: 409–419
- 16 Wang C H, Yang C T, Moreau D, et al. Motor expertise modulates neural oscillations and temporal dynamics of cognitive control. *NeuroImage*, 2017, 158: 260–270
- 17 Wu Y, Leliveld M C, Zhou X. Social distance modulates recipient's fairness consideration in the dictator game: An ERP study. *Biol Psychol*, 2011, 88: 253–262
- 18 Holroyd C B, Coles M G H. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*, 2002, 109: 679–709
- 19 Holroyd C B, Krigolson O E, Lee S. Reward positivity elicited by predictive cues. *Neuroreport*, 2011, 22: 249–252
- 20 Nieuwenhuis S, Holroyd C B, Mol N, et al. Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neurosci Biobehav Rev*, 2004, 28: 441–448
- 21 Li D Y, Li P, Li H. The updated theories of feedback-related negativity in the last decade (in Chinese). *Adv Psychol Sci*, 2018, 26: 1642–1650 [李丹阳, 李鹏, 李红. 反馈负波及其近1610年理论解释. 心理科学进展, 2018, 26: 1642–1650]
- 22 Sambrook T D, Goslin J. A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychol Bull*, 2015, 141: 213–235
- 23 Gehring W J, Willoughby A R. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science*, 2002, 295: 2279–2282
- 24 Walsh M M, Anderson J R. Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neurosci Biobehav Rev*, 2012, 36: 1870–1884
- 25 Bismark A W, Hajcak G, Whitworth N M, et al. The role of outcome expectations in the generation of the feedback-related negativity. *Psychophysiology*, 2013, 50: 125–133
- 26 Holroyd C B, Umemoto A. The research domain criteria framework: The case for anterior cingulate cortex. *Neurosci Biobehav Rev*, 2016, 71: 418–443
- 27 Boksem M A S, De Cremer D. Fairness concerns predict medial frontal negativity amplitude in ultimatum bargaining. *Soc Neurosci*, 2010, 5: 118–128
- 28 Polezzi D, Daum I, Rubaltelli E, et al. Mentalizing in economic decision-making. *Behav Brain Res*, 2008, 190: 218–223

- 29 Zhang H, Ma H Y, Xu F M, et al. Fairness preferences in the ultimate game: A dual-system theory perspective (in Chinese). *Adv Psychol Sci*, 2018, 26: 319–330 [张慧, 马红宇, 徐富明, 等. 最后通牒博弈中的公平偏好: 基于双系统理论的视角. *心理科学进展*, 2018, 26: 319–330]
- 30 Ma Q, Hu Y, Jiang S, et al. The undermining effect of facial attractiveness on brain responses to fairness in the ultimatum game: An ERP study. *Front Neurosci*, 2015, 9: 77
- 31 Qu C, Wang Y, Huang Y. Social exclusion modulates fairness consideration in the ultimatum game: An ERP study. *Front Hum Neurosci*, 2013, 7: 505
- 32 Osinsky R, Mussel P, Ohrlein L, et al. A neural signature of the creation of social evaluation. *Soc Cogn Affect Neurosci*, 2014, 9: 731–736
- 33 Gray H M, Ambady N, Lowenthal W T, et al. P300 as an index of attention to self-relevant stimuli. *J Exp Soc Psychol*, 2004, 40: 216–224
- 34 Leng Y, Zhou X. Modulation of the brain activity in outcome evaluation by interpersonal relationship: An ERP study. *Neuropsychologia*, 2010, 48: 448–455
- 35 Nieuwenhuis S, Aston-Jones G, Cohen J D. Decision making, the P3, and the locus coeruleus—Norepinephrine system. *Psychol Bull*, 2005, 131: 510–532
- 36 Yu R, Hu P, Zhang P. Social distance and anonymity modulate fairness consideration: An ERP study. *Sci Rep*, 2015, 5: 13452
- 37 Sato A, Yasuda A, Ohira H, et al. Effects of value and reward magnitude on feedback negativity and P300. *Neuroreport*, 2005, 16: 407–411
- 38 Wu Y, Zhou X. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res*, 2009, 1286: 114–122
- 39 Gu R, Ge Y, Jiang Y, et al. Anxiety and outcome evaluation: The good, the bad and the ambiguous. *Biol Psychol*, 2010, 85: 200–206
- 40 Hajcak G, Holroyd C B, Moser J S, et al. Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, 2005, 42: 161–170
- 41 Moser A, Gaertig C, Ruz M. Social information and personal interests modulate neural activity during economic decision-making. *Front Hum Neurosci*, 2014, 8: 31
- 42 Hu X, Xu Z, Li Y, et al. The impact of trust decision-making on outcome processing: Evidence from brain potentials and neural oscillations. *Neuropsychologia*, 2018, 119: 136–144
- 43 Crockett M J, Siegel J Z, Kurth-Nelson Z, et al. Moral transgressions corrupt neural representations of value. *Nat Neurosci*, 2017, 20: 879–885
- 44 Crockett M J, Kurth-Nelson Z, Siegel J Z, et al. Harm to others outweighs harm to self in moral decision making. *Proc Natl Acad Sci USA*, 2014, 111: 17320–17325
- 45 Moll J, Krueger F, Zahn R, et al. Human fronto-mesolimbic networks guide decisions about charitable donation. *Proc Natl Acad Sci USA*, 2006, 103: 15623–15628
- 46 Delorme A, Makeig S. Eeglab: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, 2004, 134: 9–21
- 47 Keil A, Debener S, Gratton G, et al. Committee report: Publication guidelines and recommendations for studies using electroencephalography and magnetoencephalography. *Psychophysiology*, 2014, 51: 1–21
- 48 Luck S J, Gaspelin N. How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, 2017, 54: 146–157
- 49 Kaltwasser L, Hildebrandt A, Wilhelm O, et al. Behavioral and neuronal determinants of negative reciprocity in the ultimatum game. *Soc Cogn Affect Neurosci*, 2016, 11: 1608–1617
- 50 Nichols S. Norms with feeling: Towards a psychological account of moral judgment. *Cognition*, 2002, 84: 221–236
- 51 Moll J, de Oliveira-Souza R, Bramati I E, et al. Functional networks in emotional moral and nonmoral social judgments. *NeuroImage*, 2002, 16: 696–703
- 52 Moll J, Eslinger P J, Oliveira-Souza R. Frontopolar and anterior temporal cortex activation in a moral judgment task: Preliminary functional MRI results in normal subjects. *Arq Neuro-Psiquiatr*, 2001, 59: 657–664
- 53 Frank R H. *Passions Within Reason: The Strategic Role of the Emotions*. New York: WW Norton & Co, 1988
- 54 Nature G. *The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard UP, 1996
- 55 DeScioli P. The side-taking hypothesis for moral judgment. *Curr Opin Psychol*, 2016, 7: 23–27
- 56 DeScioli P, Kurzban R. A solution to the mysteries of morality. *Psychol Bull*, 2013, 139: 477–496
- 57 Haidt J. The emotional dog and its rational tail: A social intuitionist approach to moral judgment. *Psychol Rev*, 2001, 108: 814–834
- 58 Greene J, Haidt J. How (and where) does moral judgment work? *Trends Cogn Sci*, 2002, 6: 517–523
- 59 Horberg E J, Oveis C, Keltner D, et al. Disgust and the moralization of purity. *J Pers Soc Psychol*, 2009, 97: 963–976
- 60 Pizarro D, Inbar Y, Helion C. On disgust and moral judgment. *Emotion Rev*, 2011, 3: 267–268
- 61 Bonini N, Hadjichristidis C, Mazzocco K, et al. Pecunia olet: The role of incidental disgust in the ultimatum game. *Emotion*, 2011, 11: 965–969
- 62 Moretti L, di Pellegrino G. Disgust selectively modulates reciprocal fairness in economic interactions. *Emotion*, 2010, 10: 169–180
- 63 Wang Y W, Wang Y, Lin C D, et al. Modulation of conscientiousness on medial frontal negativity in negative emotions: An ERP study on

- ultimatum game (in Chinese). *Sci China Life Sci*, 2011, 41: 320–331 [王益文, 王钰, 林崇德, 等. 内侧额叶负波受负性情绪下责任感影响: 最后通牒任务的 ERP 研究. *中国科学: 生命科学*, 2011, 41: 320–331]
- 64 DeSteno D, Bartlett M Y, Baumann J, et al. Gratitude as moral sentiment: Emotion-guided cooperation in economic exchange. *Emotion*, 2010, 10: 289–293
- 65 Wang G, Li J, Li Z, et al. Medial frontal negativity reflects advantageous inequality aversion of proposers in the ultimatum game: An ERP study. *Brain Res*, 2016, 1639: 38–46
- 66 Fehr E, Gächter S. Altruistic punishment in humans. *Nature*, 2002, 415: 137–140
- 67 Mendez M F. The neurobiology of moral behavior: Review and neuropsychiatric implications. *CNS Spectr*, 2009, 14: 608–620
- 68 van Nunspeet F, Ellemers N, Derks B, et al. Moral concerns increase attention and response monitoring during IAT performance: ERP evidence. *Soc Cogn Affect Neurosci*, 2014, 9: 141–149
- 69 Chang L J, Sanfey A G. Great expectations: Neural computations underlying the use of social norms in decision-making. *Soc Cogn Affect Neurosci*, 2013, 8: 277–284
- 70 Martin R S. Event-related potential studies of outcome processing and feedback-guided learning. *Front Hum Neurosci*, 2012, 6: 304
- 71 Qu C, Métiveau E, Butera L, et al. Neurocomputational mechanisms at play when weighing concerns for extrinsic rewards, moral values, and social image. *PLoS Biol*, 2019, 17: e3000283
- 72 Spitzer M, Fischbacher U, Herrnberger B, et al. The neural signature of social norm compliance. *Neuron*, 2007, 56: 185–196
- 73 Bénabou R, Tirole J. Incentives and prosocial behavior. *Am Econ Rev*, 2006, 96: 1652–1678
- 74 Linden D E J. The P300: Where in the brain is it produced and what does it tell us? *Neuroscientist*, 2005, 11: 563–576
- 75 Cuthbert B N, Schupp H T, Bradley M M, et al. Brain potentials in affective picture processing: Covariation with autonomic arousal and affective report. *Biol Psychol*, 2000, 52: 95–111
- 76 Pollatos O, Kirsch W, Schandry R. On the relationship between interoceptive awareness, emotional experience, and brain processes. *Cogn Brain Res*, 2005, 25: 948–962
- 77 Decety J, Porges E C. Imagining being the agent of actions that carry different moral consequences: An fMRI study. *Neuropsychologia*, 2011, 49: 2994–3001
- 78 Hauser T U, Iannaccone R, Stämpfli P, et al. The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. *NeuroImage*, 2014, 84: 159–168

Summary for “道德判断影响大脑对公平行为的早期评价”

Moral judgment modulates fairness consideration in the early outcome evaluation stage

Bin Zhan[†], Bixuan Du[†], Shaohua Chen, Yiwen Li, Weiqi He^{*} & Wenbo Luo^{*}

Research Center of Brain and Cognitive Neuroscience, Liaoning Normal University, Dalian 116029, China

[†] Equally contributed to this work

^{*} Corresponding authors, E-mail: hewq@lnnu.edu.cn; luowb@lnnu.edu.cn

Morality, concern for the welfare of others, is an important criterion of how individuals treat others, including justice, fairness, and rights. Moral judgment is the process of evaluating the behavior of others and oneself according to social norms. Morality might affect social cooperation with others by entailing the belief that someone should be rewarded or punished. The ultimatum game is a classic paradigm that can be used to study the fairness behavior of social cooperation. Previous studies have shown that some social factors (e.g., social distance, social exclusion, and social evaluation) influence fairness behavior in the ultimatum game. Using the event-related potential (ERP) technique to examine the neural basis of the ultimatum game, studies found that feedback-related negativity (FRN) was modulated by social factors. This effect indicates that the influence of social factors mainly appears in the early stage of outcome evaluation. However, the influence of moral judgment on social cooperation and its neural mechanism remains unclear. The present study focused on whether moral judgment would affect fairness consideration in the early outcome evaluation stage.

For this purpose, we used moral sentences within the ultimatum game to assess the influence of moral judgment and fairness of offers on decision-making. A total of 25 participants who acted as respondents completed the ultimatum game with blood donors, murderers, and unidentified individuals as proposers. A moral sentence (e.g., He is a blood donor) was given as a reminder of the moral identity of the players before each round. Participants were told that other players would gain ¥50, and proposers would make an allocation offer. There were five types of allocation offers: ¥5, ¥10, ¥15, ¥20 or ¥25. Participants were asked to accept or reject the offers. At the same time, the ERP was locked to the onset of the offers. The acceptance rate of each offer was included in the behavioral data. The dependent variable was the acceptance rate of the five allocation offers and moral primes were the independent variable when performing one-way ANOVAs. Regarding ERP data, ¥5 and ¥10 were considered as unfair offers and ¥20 and ¥25 as fair offers in order to analyze the amplitudes of FRN and P300. Two-way repeated measures ANOVAs (moral judgment, three levels; offer fairness, two levels) were performed to compare ERP data.

Behavioral results showed that the acceptance rate of unfair offers from blood donors was significantly higher than that of murderers. Furthermore, the acceptance rate of fair offers from murderers was significantly lower than that of blood donors and unidentified individuals. ERP results revealed that the FRN amplitude was more negative-going in response to unfair than to fair offers. The fairness of offers modulated FRN amplitude only regarding blood donors, but not for murderers and unidentified individual conditions. P300 was larger for fair offers compared to unfair offers and was not affected by moral judgment. These results indicate that when moral judgment conflicts with the fairness criterion, individuals might rely on moral judgment to make decisions. People may have a higher expectation of fairness from individuals with high moral standards, while unfair behaviors could induce strong negative emotions. FRN represents unfair behaviors that violate social norms and shows that the early outcome evaluation stage is also modulated by moral judgment, while the late stage is independent of moral judgment.

moral judgment, ultimatum game, feedback-related negativity (FRN), P300, fairness consideration

doi: [10.1360/TB-2019-0681](https://doi.org/10.1360/TB-2019-0681)