

社会决策中的公平准则及其神经机制*

罗 艺¹ 封春亮¹ 古若雷² 吴婷婷¹ 罗跃嘉^{1,3}

(¹ 北京师范大学认知神经科学与学习国家重点实验室, 北京 100875)

(² 中国科学院心理研究所, 北京 100101) (³ 成都医学院, 成都 610083)

摘 要 社会决策中的公平准则一直受到广泛的关注, 经济学家和心理学家们分别从动机和认知的角度对这个课题进行了深入探索。在这些工作中, 最为突出的是“均等”、“互惠”和“情绪”三个模型的提出。均等和互惠模型试图从动机角度理解个体的公平行为, 而情绪模型则更强调个体在公平决策中的认知和情绪的加工过程。近年来脑成像的研究也为这些模型提供了相应的支持, 其中与公平准则最密切的脑区是脑岛和背外侧前额叶: 前者一般被认为是反映了个体面对不公平待遇时的负性情绪体验; 而后者主要负责决策过程中的抑制控制。但这些脑区具体的功能还存在很多争议, 反映了不同理论模型对特定脑区功能的不同理解。未来研究应注意三个模型之间的相互补充, 以及加强多模态的脑指标的运用, 并注重相关脑区的整体性, 此外, 公平决策行为中的个体差异, 以及社会情境对其的影响也应受到重视。

关键词 公平准则; 社会决策; 情绪; 脑岛; 背外侧前额叶

分类号 B842

1 公平规则对社会决策的影响

古代思想家很早就认识到公平在社会活动中的重要性, 比如儒家经典《论语》就提出了“不患贫而患不均”的观点——治国者需要担心的不是社会财富总量不足, 而是财富在人民之中分配不平等。北宋理学家陈襄也曾指出: 为官者“事惟公平, 可以服人心”(引自《州县提纲》)。后世的学者也一直延续着“什么是公平”的思考(Harlé, Chang, van 't Wout, & Sanfey, 2012; Picton, Saunders, & Jentsch, 2012; Suveg & Zeman, 2004; Wang, Seidler, Hall, & Preston, 2012), 充分表明公平这个主题在人类社会活动中占有重要地位。到了现代, 对公平的探讨已经不再是哲学家的“一言堂”。心理学家和经济学家都积极参与其中, 对公平相关的心理机制、脑机制以及进化机制进行

了深入研究(Momennejad & Haynes, 2012)。下文将具体介绍对人类执行与遵守公平准则的倾向进行解释的三种主流理论模型, 以及相关的实验证据。

1.1 均等模型

古典经济学理论认为, 最优的决策策略就是使个体利益最大化的策略。为了做出最优选择, 个体必须遵循“理性人”的思维模式, 避免受到情绪或其他社会因素的影响(Sanfey, Rilling, Aronson, Nystrom, & Cohen, 2003)。然而研究者们很早就发现在社会决策情景中, 除了单纯的利益导向之外, 公平也是影响个体行为的一个重要因素(Rabin, 1993)。个体执行与遵守公平准则的倾向, 在使用“最后通牒游戏(Ultimatum Game, UG)”范式的实验中得到了充分体现(Güth, Schmittberger, & Schwarze, 1982)。UG 游戏一般包括两名参加者, 分别扮演“分配者”和“接收者”两种角色, 对一笔奖金进行分配。其中分配者有权提出分配方案, 接收者有权接受或拒绝该方案。如果分配者的方案被接受, 这笔奖金就按照分配者提出的方案进行分配; 如果方案被拒绝, 两人的所得金额均为零。按照“理性人”假说, 基于利益最大化的原则,

收稿日期: 2012-07-16

* 国家自然科学基金重大研究计划培育项目(91132704), 国家自然科学基金重点项目(30930031), 国家 973 计划项目(2011CB711000), 科技部支撑计划课题(2009BAI77B01)。

通讯作者: 罗跃嘉, E-mail: luoyj@bnu.edu.cn

分配者应该会分尽可能少的金额给对方，而接收者应该接受任何非零的分配方案，因为任何数额的收入都比得不到钱更为有利。然而大量实验结果表明分配者往往倾向于将一半左右的奖金分给对方；而当分配给接收者的份额少于20%时，这些分配方案被拒绝的概率高达50% (Sanfey et al., 2003; Schwartz & Weinberger, 1980)。

在Fehr与Schmidt (1999)提出的均等模型中，个体的拒绝行为被解释为对不均等结果的厌恶 (Chen & Vazsonyi, 2011)。具体来说，个体在社会决策中不仅仅会考虑自身的所得，同时也会考虑自己与别人的收入是否对等 (Tabibnia, Satpute, & Lieberman, 2008)。而当个体与他人的所得处于不均等状态时，那么即使个体的所得不为零，个体的心理收益也可能是负值，从而促使个体做出拒绝行为以达到均等的目的。这个均等模型得到了大量实验结果的支持。例如，Fehr等(2008)发现儿童随着年龄的增长，在对资源(糖果)进行分配时会越来越多地倾向于均等方案，回避不均等方案(不论方案对自己有利还是不利)，这说明公平的观念在儿童社会化的过程中得到了强化。为了消除自己与他在分配上的不均等状态，UG中的接收者会牺牲自身利益来拒绝不公平的分配方案，从而对分配者造成惩罚(Dawes, Fowler, Johnson, McElreath, & Smirnov, 2007; Dawes et al., 2012; Johnson, Dawes, Fowler, McElreath R, & Smirnov, 2009)。最后，近年来脑成像的研究也发现均等的分配会激活奖赏加工的脑区(Harlé et al., 2012; Tabibnia et al., 2008; Tricomi, Rangel, Camerer, & O'Doherty, 2010)，表明均等状态本身对个体来说就是一种奖赏。

虽然已有大量的研究支持均等模型，但也有一些实验结果与均等模型并不相容。例如，与电脑随机提出的分配方案相比，接收者更倾向于拒绝由人提出的方案，即使两者的不均等程度是一样的(Blount, 1995; Knoch et al., 2008; Knoch, Pascual-Leone, Meyer, Treyer, & Fehr, 2006)。但是根据均等模型，被试的公平感应该只取决于分配方案是否均等，而不管方案的提出者是电脑还是人。因此分配者的变化引起的行为差异无法用厌恶不均等的动机来解释。

1.2 互惠模型

另一种理论模型——互惠(reciprocity)——

能更好地解释为什么个体更倾向于拒绝由人提出的分配方案。研究者们认为，存在三种互惠原则：第一种是直接互惠(direct reciprocity)，即个体当前的行为能直接给自己带来利益。例如当个体之间需要多次重复交往的时候，当前的惩罚行为可能是为了在后续的交往中获得更多的利益 (Trivers, 1971)；第二种是间接互惠(indirect reciprocity)，即个体虽然不能通过惩罚从互动的对方那里直接得到利益，但却可以藉此在群体中建立声誉，从而在之后与他人的社会交往中得到更多公平对待，间接地获益(Nowak, 2006)。然而不管是直接互惠还是间接互惠，都不能很好地解释在UG中出现的拒绝行为，因为UG的分配者和接受者之间的互动往往是单次(one-shot)的而不是多次重复的，而且也没有第三方可以获得接受者是否拒绝(亦即给予惩罚)的信息。因此个体的惩罚行为可能反映了第三种互惠原则：强互惠(strong reciprocity)，也即个体出于维护公平准则的动机而采取惩罚行为，即使惩罚行为本身是高成本的，而且惩罚实施者并不能直接或间接地从中获得任何利益(Bowles & Gintis, 2004)。

根据强互惠原则，UG中的惩罚行为的性质是利他的。虽然利他惩罚对个体没有意义，但如果一个群体中存在强互惠的个体，这样的群体在进化中是有优势的：强互惠者(strong reciprocator)会惩罚违反公平准则的行为，使公平准则在群体内得到巩固(Bruni, Gilli, & Pelligra, 2008)。因此互惠模型能更好地解释为什么UG接收者更倾向于拒绝由人提出的不均等分配方案；而当分配者是电脑时，这种互惠的效应就消失了。

均等模型和互惠模型，都是从宏观的动机角度去解释公平行为的产生。然而随着对公平研究的日渐深入，这两种经典模型还是暴露出了一定的局限性。例如，在Yamagishi等人(2009)设计的“免责游戏(Impunity Game)”中，接收者的决策只会影响他自己的所得；无论接收者选择接受或拒绝，分配者都可以得到他分给自己的那部分金额。即使在这样的实验情景下，接收者拒绝不公平方案的可能性仍然高达30%—40%。均等和互惠模型都不能解释接收者在免责游戏中的追求公平的倾向：一方面，根据游戏规则，接收者的拒绝行为实际上会增大两位参与者之间的分配不均等程度；另一方面，接收者的拒绝行为无法影响分

配者的所得。分配者甚至不会知道对方是否拒绝了自己的分配方案,因此免责游戏的接收者是无法惩罚不公平分配者,从而让对方遵守公平准则的。可见,仅仅从动机的角度解释个体在社会决策中的公平行为,是存在很大局限性的。研究者们发现,如果在理论模型中引入情绪因素,就可以对 Yamagishi 等人的发现做出相对合理的解释:不公平的分配方案会诱发个体的负性情绪,而拒绝不公平的行为,可能是对这种负性情绪的表达和宣泄(Xiao & Houser, 2005)。

1.3 情绪模型

情绪模型认为,个体的情绪因素在社会决策中起至关重要的作用(Pillutla & Murnighan, 1996)。以往很多研究已经发现不公平的分配方案会引起负性情绪反应(Calder, Keane, Manes, Antoun, & Young, 2000; Damasio et al., 2000; Hewig et al., 2011; van't Wout, Kahn, Sanfey, & Aleman, 2006),为我们深入理解情绪对公平行为的推动作用提供了重要信息。为了验证情绪模型,研究者们试图在UG中操纵个体的情绪状态,以考察个体在不同情绪状态下面对不公平分配时的行为变化。例如在Harlé和Sanfey(2007)的研究中,通过电影片段诱发被试的悲伤、喜悦或中性情绪,结果发现悲伤情绪组比中性情绪组拒绝了更多的不公平分配方案。

另一方面,情绪调节也能显著改变个体在UG中的行为模式。Xiao等人在经典UG中加入了一个书写表达阶段——接受者在做出选择之前,可以在主试提供的卡片上向分配者写下一句话。他们发现与控制组(无书写表达)相比,书写表达组的接受率出现显著提高。这说明UG中的拒绝行为是接收者对不公平分配表达愤怒的一种手段,而书写表达使接收者在决策前宣泄了愤怒,从而减弱了他们接下来执行惩罚的意愿。最近一项研究比较了不同类型的情绪调节策略对UG的影响,结果发现与“表达抑制”策略相比,“认知重评”策略对UG中的决策倾向影响更大,会使接收者更多地接受不公平的分配方案(van't Wout, Chang, & Sanfey, 2010)。Grecucci等人(2012)进一步对比了两种认知重评策略——向上调节(up-regulation, 增强对分配者意图的负性评价)和向下调节(down-regulation, 减弱负性评价)——对个体在UG中的行为和大脑活动的影响,发现向上调节

使拒绝率提高,而向下调节的效果则相反。

最后,一项针对脑损伤病人的研究发现,与正常人或者其他脑区受损的病人相比,腹内侧前额叶(ventral medial prefrontal cortex, VMPFC)损伤病人对不公平分配方案的拒绝率显著提高。VMPFC被认为在情绪调节中起重要作用,因此这类病人在UG中的高拒绝率可能是情绪调节能力受损所导致的(Koenigs & Tranel, 2007)。这项研究从侧面证明,不公平待遇会引起愤怒、厌恶等负性情绪体验,而个体对这些负性情绪的有效调节则能显著地降低惩罚的动机。

近年来流行的“双系统理论”(dual system approach)试图融合社会决策中的认知因素和情绪因素。这个理论认为,在社会决策过程中存在两个系统的交互作用,一个是反应较快的、自下而上的情绪系统,与自动化的过程相关;另一个则是反应相对较慢的、自上而下的理性系统,与抑制控制相关(Loewenstein & O'Donoghue, 2004)。实证研究发现被试拒绝分配方案的决策用时显著短于接受方案的用时;此外,当个体处于高认知负荷状态下时,对不公平方案的拒绝率会显著提高。这些结果说明拒绝不公平方案的行为属于较快的、自动化的反应;而接受不公平方案的倾向会受到认知资源的影响,可能与理性系统的抑制控制过程有关(Halali, Bereby-Meyer, & Meiran, 2011)。也就是说,公平在社会决策中的作用机制并不是简单地受情绪所驱动,而是一个认知和情绪交互作用的复杂过程。

以上三个理论模式是公平决策领域最受关注的模型,很多研究都是围绕这三个理论模型的争论展开(Xiao & Houser, 2005; Yamagishi et al., 2009),并且,在Rilling和Sanfey(2011)的综述中,这三个模型的解释力得到了总结和肯定。以往研究中的争论主要来源于这些理论的不同预测。这些争论不仅仅局限于行为层面,同时也包括这些模型对公平决策相关脑机制所作出的不同解释。

2 公平规则影响社会决策的脑机制

近十年来随着脑成像技术的日趋成熟,公平行为所对应的脑机制也逐步被揭示出来。研究者们认为,在与公平感相关的脑区之中,最重要的两个区域是主要负责情绪加工的脑岛(insula)和主要负责认知控制的背外侧前额叶(dorsal lateral

prefrontal cortex, DLPFC)。

2.1 脑岛

脑岛可以分成多个功能不同的亚区, 其中后脑岛主要负责编码初级的数量信息(Wright, Symmonds, Fleming, & Dolan, 2011); 中脑岛主要负责将身体的生理状态或感觉, 与意识相关的线索相整合(Craig, 2009; Wright et al., 2011); 前脑岛主要涉及对情绪和身体状态的自我意识(Paulus & Stein, 2006; Singer, Critchley, & Preuschoff, 2009)。在UG任务中, 不公平的分配方案主要激活的是前脑岛, 而且前脑岛的激活强度与对不公平方案的拒绝率呈正相关(Sanfey et al., 2003)。由于前脑岛在物理性和社会性的负性情绪体验中都扮演重要角色(Harenski & Hamann, 2006; Moll et al., 2002; Singer et al., 2004), 研究者们将前脑岛在UG任务中的激活, 解释为个体在面对不公平分配方案时的厌恶或愤怒情绪体验。这也从侧面支持了情绪模型的假说(Güroğlu, van den Bos, Rombouts, & Crone, 2010; Güroğlu, van den Bos, van Dijk, Rombouts, & Crone, 2011; Rilling & Sanfey, 2011; Sanfey, 2007; Tabibnia et al., 2008; Wright et al., 2011)。近年来, 研究者通过操纵个体在进行UG任务时的情绪状态, 进一步证明前脑岛的激活确实反映了面对不公平方案时的负性情绪反应(Grecucci, Giorgetta, Van't Wout, Bonini, & Sanfey, 2012; Harlé et al., 2012)。例如, 与诱发中性情绪的个体相比, 被诱发悲伤情绪的个体在面对不公平分配时出现了更强的前脑岛的激活, 对这类方案的拒绝率也相应提高(Harlé et al., 2012)。这证明情绪因素在公平相关的社会决策中起着重要的作用, 而前脑岛则是其中的关键脑区。

近期的研究发现后脑岛在公平相关的决策中也会被激活。有人认为, 它主要编码的是分配方案在分配者和接受者之间的均等程度, 而对分配方案的社会情境不敏感; 与之相反, 前脑岛不仅编码分配方案的均等性, 同时也能反映出社会情境的微妙差异(Wright et al., 2011), 提示前脑岛的活动能更加灵活地反映个体在特定社会情境下的情绪状态, 这和Harlé等(2012)的结论是一致的。但是在—项情绪调节研究中, Grecucci等(2012)发现后脑岛在面对不公平分配时的激活能够反映出不同调节策略(向上或向下调节)之间的差异, 表明后脑岛也可能对不同情景下的情绪状态敏感,

而不仅仅反映了分配方案本身是否均等。总括而言, 虽然大多数研究者都同意脑岛的活动反映了个体面对不公平分配时的负性情绪体验, 但脑岛不同亚区在公平相关决策中的具体作用还有待进一步的研究。

2.2 DLPFC

对DLPFC功能的一种较普遍的解释是认知控制假设(cognitive control hypothesis)。根据这个假设, DLPFC在社会决策中的激活反映的是认知控制以及对优势反应的抑制。然而对UG接收者来说, 拒绝(“非理性”/利他)和接受(“理性”/利己)这两种选择, 哪一种才是优势反应? 关于这一点尚存在激烈的争论。正如前文所述, 双系统理论认为, 对不公平分配的厌恶情绪会促使个体产生拒绝分配的冲动, 这种反应是较快的、自下而上的。因此, 拒绝不公平的方案是个体的优势反应, 而DLPFC抑制的就是个体的负性情绪及相应的拒绝反应, 从而让个体做出更符合“理性”的选择(Sanfey et al., 2003)。但是Knoch等人(2006)使用经颅磁刺激技术(Transcranial Magnetic Stimulation, TMS), 发现右侧DLPFC受到刺激干扰的被试组对人提出的不公平方案的接受率显著高于控制组。这项研究不支持DLPFC抑制负性情绪的观点, 因为如果DLPFC抑制的是不公平方案带来的负性情绪, 那么当该抑制功能被削弱(DLPFC活动受抑制)之后, 负性情绪对决策的影响应该更大, 也就是说被试会做出更多的拒绝行为, 而这与Knoch等人的实验发现恰恰相反。因此, Knoch等人认为个体具有自我获利的冲动——在面对不公平分配方案时, 个体的优势反应是接受方案, 因为这能直接使自己获利, 而DLPFC则在决策过程中抑制这种冲动, 使个体更多地考虑公平准则, 通过拒绝不公平方案来惩罚分配者。我们可以看出, 对DLPFC功能的争论其实反映了情绪理论与互惠理论对UG中的拒绝行为的不同理解: 情绪理论认为拒绝行为是一种情绪化的非理性行为, 当个体面对不公平方案时, 需要DLPFC来抑制自己的非理性冲动; 互惠理论则认为拒绝行为是出于利他的动机, 因此个体做出拒绝行为时需要DLPFC抑制利己的倾向。

近期, 一些研究者对DLPFC的作用给出了另一种解释——整合选择假设(integration-and-selection hypothesis)。这个假设认为, 在社会决策

的过程中, DLPFC 可能负责表征长远的、全局性的目标和奖赏, 这些表征以及所有的选项和相应后果被传达到背侧纹状体等其他脑区。这些脑区的协同工作, 使个体能够根据与情景相符的规则, 整合双方(或多方)的责任、自己行为可能造成的伤害等各方面信息, 从而做出在当前社会情境下最适合的行为。当 DLPFC 的活动受到抑制时, 个体也就失去了对长远利益的表征能力, 仅仅依据当前获利指导自己的行为。在这种情况下, 对不公平方案的拒绝率会出现下降(Buckholz & Marois, 2012)。整合选择假设还能解释认知控制假设所无能为力的一些现象, 例如 Buckholz 等人(2008)发现, 与不惩罚的选择相比, 当个体在第三方惩罚任务中选择惩罚时, DLPFC 的活动更强。这是因为在选择惩罚时, 个体需要整合各方面的信息; 而不选择惩罚时, 个体只需要简单地抑制自己的反应。因此这个结果提示 DLPFC 主要负责对各类信息的整合, 而不是简单的抑制(Knoch et al., 2006, 2008; Sanfey et al., 2003)。

在 DLPFC 功能的偏侧化方面, 现有的研究结果也存在一些矛盾。早期的研究发现, 不公平方案能比公平方案诱发更强的双侧 DLPFC 激活(Sanfey et al., 2003)。后续的研究使用经颅磁刺激(TMS)和经颅电刺激(transcranial direct current stimulation, tDCS)方法分别干扰左右侧 DLPFC, 结果发现只有在右侧 DLPFC 被抑制时, 维护公平准则的行为才显著减少(Baumgartner, Knoch, Hotz, Eisenegger, & Fehr, 2011; Knoch et al., 2006, 2008), 提示只有右侧 DLPFC 在公平相关行为中起作用。近年来的一些发展心理学研究则对这个结论提出了挑战(Figner et al., 2010; Hare, Camerer, & Rangel, 2009), 如 Harlé 等人(2012)发现相比于年轻人来说, 老年人在参与 UG 时更依赖于左侧 DLPFC 的激活来维持抑制控制等功能。此外, Steinbeis 等人(2012)以青少年和成年人为被试的研究发现, 随着个体年龄的增长, 左侧 DLPFC 的激活与对公平准则的遵守呈正相关, 也与冲突控制的能力呈正相关。在排除年龄因素之后, 左侧 DLPFC 的皮层厚度仍然可以预测个体对公平准则的遵守倾向。这些互相冲突的结果可能说明左侧和右侧 DLPFC 在个体执行与遵守公平准则时都起着一定的作用, 将来的研究则需要进一步分离左侧和右侧 DLPFC 在公平相关决策中所对应

的心理机制。

2.3 其他脑区

综合上文可知, 在社会决策过程中, 前脑岛激活反映的可能是不公平分配诱发的负性情绪体验, 而 DLPFC 激活反映的可能是对这种情绪的抑制控制。除了这两个被重点强调的区域之外, 研究者们也逐步发现了其他脑区在公平行为中的作用。例如, 公平行为涉及的情绪加工脑区可能并不局限于前脑岛。近期研究表明, 不公平的分配方案能引起杏仁核(amygdala)的激活, 而且其活动强度能预测个体对不公平方案的拒绝率(Gospic et al., 2011)。考虑到杏仁核在多种基本情绪(例如恐惧和焦虑)加工中的重要作用(Feinstein, Adolphs, Damasio, & Tranel, 2011; Hamann, 2011), 这项研究的结果应该说是与情绪模型的预期一致的。但是杏仁核与脑岛的关系, 在很大程度上还是目前研究的一个空白——在公平相关决策中, 两个脑区是协同编码个体的情绪体验, 还是有所分工, 各司其职? 这个问题只能等待后续实验的进一步挖掘。

另一方面, 在公平行为中参与认知控制的脑区也不仅仅局限于前额叶。例如在 UG 中, 不公平方案与公平方案相比, 前者诱发的扣带前回(anterior cingulate cortex, ACC)激活更强(Güroğlu et al., 2010; Sanfey et al., 2003)。ACC 被普遍认为与冲突监控有关(Botvinick, Nystrom, Fissell, Carter, & Cohen, 1999; MacDonald, Cohen, Stenger, & Carter, 2000), 因此 ACC 在 UG 中的激活模式可能反映了双系统理论中所强调的认知与情绪动机的冲突。再者, 腹外侧前额叶(ventrolateral prefrontal cortex, VLPFC)等参与情绪调节的脑区也在公平相关决策行为中起着重要作用——当不公平方案被接受时, 往往伴随着 VLPFC 等情绪调节相关脑区活动增强, 以及前脑岛活动削弱(Hariri, Mattay, Tessitore, Fera, & Weinberger, 2003; Lieberman, 2011; Tabibnia et al., 2008)。

3 总结和展望

3.1 三个模型的相互补充

现代经济学家和心理学家从动机、情绪等角度探讨了个体对公平准则的执行与遵守, 提出了均等、互惠、情绪等理论模型来解释这种人类特有的社会行为。通过将经典范式与脑成像技术相

结合, 研究者们可以探索脑岛、DLPFC 等脑区在社会决策过程中的功能, 使我们对个体执行与遵守公平准则所涉及的神经机制的认识越来越深入。然而现有的几个理论模型之间存在冲突, 对个体拒绝不公平方案的行为所作出的解释很不一致。例如, 均等模型认为个体的拒绝行为仅仅是为了消除与他人之间不均等的收益状态, 也即一种类似于“仇富”的动机(Dawes et al., 2012; Johnson et al., 2009); 互惠模型认为个体的拒绝行为是出于一种利他的动机, 是为了维护“公平”这个社会准则(Bruni et al., 2008); 情绪模型则强调不公平方案诱发的负性情绪是个体做出拒绝行为的主要因素, 不管个体的动机是利他的还是非利他的(Harlé et al., 2012)。这三种理论模型从不同角度寻找个体执行与遵守公平准则的原因, 这其实也反映了个体社会行为客观上的复杂性。虽然它们都有相应的实验证据支持, 但迄今甚少有研究探讨不同模型之间的联系。未来研究应注重探索三种理论模型是否能够相互补充, 如个体在面对不公平方案时, 情绪因素的产生与作用, 可能受到互惠模型中所强调的个体自身的道德标准, 或者说个体的利他动机的影响, 由此, 情绪模型应该参考互惠模型的内容, 进行一定的修正。

3.2 公平决策行为中的个体差异

考察公平决策行为中的个体差异, 也是一个值得关注的未来研究方向。Fehr 和 Schmit (1999) 提出的均等模型已经考虑到了个体差异的因素, 认为不同个体对分配不均等的敏感程度不同。因此分配方案的公平程度在不同个体的社会决策中所占的权重是不一样的。但是现有大多数研究侧重在整体样本中考察对公平准则执行与遵守的心理或神经机制, 个体间的差异则往往被忽视。具体来说, 为什么有些个体面对不公平的方案时会产生强烈的负性情绪从而拒绝这些方案, 而其他个体却能够接受非常不公平的方案? 对这个问题的回答一方面能够帮助我们更好地理解个体执行与遵守公平准则时涉及的具体动机和心理过程, 另一方面也可以检验现有理论模型中的假设。例如, 通过比较情绪调节能力不同的个体在公平相关决策中的行为以及神经活动模式差异, 可以更有效地对情绪模型进行检验。

3.3 不同模态的脑指标的运用

除了传统的问卷研究和临床精神病学研究之

外, 近几十年来神经影像学新技术的飞速发展, 也为我们从脑结构的角度考察个体差异提供了方便。上文已经提到, 大量的功能磁共振成像研究从整体水平揭示了个体执行与遵守公平准则所涉及的神经机制(Fehr et al., 2008; Harlé et al., 2012), 而近年发展起来的弥散张量成像、近红外光学成像等新技术还尚未成熟运用到社会决策领域的研究中。近年来兴起的一个热门研究领域, 就是用不同模态的脑指标来预测个体的认知能力(Lewis, Baldassarre, Committeri, Romani, & Corbetta, 2009; Zhu, Zhang, Luo, Dilks, & Liu, 2011)。这方面的现有研究还主要局限在基础认知领域(知觉、抑制控制等), 而在高级的社会认知领域(如对公平准则的执行)的应用还很少。我们认为, 社会决策中涉及的高级社会功能是基础认知活动在社会情境中的一种延伸(Buckholz & Marois, 2012), 因此在公平决策行为中体现出的个体差异, 也可能像基础认知活动一样能找到相应的神经标记。那么, 研究者就可以通过考察个体大脑的形态结构来预测他们在公平决策中的行为模式。

3.4 公平行为相关脑区的整体性

在前文中提到了许多在公平相关任务中被激活的脑区(前脑岛、DLPFC、杏仁核、ACC、VLPFC等), 而对这些脑区功能的解释还存在很多争议。一个可能的原因是, 这些脑区的功能并不是相互独立的, 而是相互影响并形成大尺度的神经网络, 因此当研究者孤立地考察这些脑区的功能时会有许多不一致的发现。可见, 后续的研究应该更加注重对公平行为相关脑区整体性的探讨, 而不是对单个脑区的活动进行解释。近年来得到广泛应用的多体素模式分析(multi-voxel pattern analysis, MVPA)技术就为这一研究思路的实现提供了可能性(Norman, Polyn, Detre, & Haxby, 2006)。与传统考察单个脑区在任务中的激活不同, MVPA 技术可以同时结合多个脑区的激活信息以及这些脑区之间的连接信息, 并且利用这些整合性的信息来预测个体的行为表现。我们认为, 研究者可以借助这项技术对公平相关的多个脑区间的联系, 以及它们与个体公平行为之间的关系做更加深入的考察。此外, 不同认知神经科学技术的结合也能够帮助我们理解脑区间的相互关系。

3.5 注重社会情境对决策过程的影响

虽然研究者们对个体在社会互动情景中的行

为非常感兴趣,但在以往的大多数研究中,个体间的互动往往被研究者严格控制。换句话说,参与者在实验室中的决策过程是在一个封闭的社会情境中进行的,虽然大多数研究者认为经典范式的这些设置避免了无关信息对个体决策的干扰,但这显然削弱了实验的生态效度。在真实生活中,个体在进行公平相关的决策时,除了考虑分配方案的公平程度之外,还需要考虑其他因素,例如与分配者的社会关系(朋友、上级等)、对方的个人信息(外貌、贫富)等。在这些不同的社会情境中,个体对公平准则的执行与遵守,可能是由不同的社会因素的影响。社会心理学家们很早就发现,个体在社会互动过程中对群体内成员和群体外成员的态度有巨大差别,即所谓的内群体偏好(Brewer, 1979)。那么个体在与群体内或群体外成员进行社会决策时,背后的行为动机可能有很大差异——拒绝群体内成员的不公平方案很可能是出于利他的动机,目的是为了通过惩罚让群体内遵守公平准则的个体增多,从而让自己所属的群体受益;而拒绝群体外成员的不公平方案可能是基于报复的动机,目的是为了对方不能得到比自己更多的收益。可见,将公平相关的决策行为置于特定的社会情境之下,可能有助于丰富现有的理论模型,让研究者对个体执行与遵守公平准则所涉及的动机以及心理机制有更为全面的理解。

参考文献

- Baumgartner, T., Knoch, D., Hotz, P., Eisenegger, C., & Fehr, E. (2011). Dorsolateral and ventromedial prefrontal cortex orchestrate normative choice. *Nature Neuroscience*, 14(11), 1468–1474.
- Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational Behavior and Human Decision Processes*, 63(2), 131–144.
- Botvinick, M., Nystrom, L. E., Fissell, K., Carter, C. S., & Cohen, J. D. (1999). Conflict monitoring versus selection-for-action in anterior cingulate cortex. *Nature*, 402(6758), 179–181.
- Bowles, S., & Gintis, H. (2004). The evolution of strong reciprocity: Cooperation in heterogeneous populations. *Theoretical Population Biology*, 65(1), 17–28.
- Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin*, 86(2), 307–324.
- Bruni, L., Gilli, M., & Pelligra, V. (2008). Reciprocity: Theory and facts. *International Review of Economics*, 55(1-2), 1–11.
- Buckholtz, J. W., Asplund, C. L., Dux, P. E., Zald, D. H., Gore, J. C., Jones, O. D., ... Marois, R. (2008). The neural correlates of third-party punishment. *Neuron*, 60(5), 930–940.
- Buckholtz, J. W., & Marois, R. (2012). The roots of modern justice: Cognitive and neural foundations of social norms and their enforcement. *Nature Neuroscience*, 15(5), 655–661.
- Calder, A. J., Keane, J., Manes, F., Antoun, N., & Young, A. W. (2000). Impaired recognition and experience of disgust following brain injury. *Nature Neuroscience*, 3, 1077–1078.
- Chen, P., & Vazsonyi, A. T. (2011). Future orientation, impulsivity, and problem behaviors: A longitudinal moderation model. *Developmental Psychology*, 47(6), 1633–1645.
- Craig, A. D. (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10, 59–70.
- Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience*, 3, 1049–1056.
- Dawes, C. T., Fowler, J. H., Johnson, T., McElreath, R., & Smirnov, O. (2007). Egalitarian motives in humans. *Nature*, 446(7137), 794–796.
- Dawes, C. T., Loewen, P. J., Schreiber, D., Simmons, A. N., Flagan, T., McElreath, R., ... Paulus, M. P. (2012). Neural basis of egalitarian behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 109(17), 6479–6483.
- Fehr, E., Bernhard, H., & Rockenbach, B. (2008). Egalitarianism in young children. *Nature*, 454(7208), 1079–1083.
- Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, 114(3), 817–868.
- Feinstein, J. S., Adolphs, R., Damasio, A., & Tranel, D. (2011). The human amygdala and the induction and experience of fear. *Current Biology*, 21(1), 34–38.
- Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., & Weber, E. U. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature Neuroscience*, 13(5), 538–539.
- Gospic, K., Mohlin, E., Fransson, P., Petrovic, P., Johannesson, M., & Ingvar, M. (2011). Limbic justice—amygdala involvement in immediate rejection in the Ultimatum Game. *Plos Biology*, 9(5), e1001054.
- Güroğlu, B., van den Bos, W., Rombouts, S. A. R. B., & Crone, E. A. (2010). Unfair? It depends: Neural correlates

- of fairness in social context. *Social Cognitive and Affective Neuroscience*, 5(4), 414–423.
- Güroğlu, B., van den Bos, W., van Dijk, E., Rombouts, S. A. R. B., & Crone, E. A. (2011). Dissociable brain networks involved in development of fairness considerations: Understanding intentionality behind unfairness. *Neuroimage*, 57, 634–641.
- Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental-analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, 3(4), 367–388.
- Grecucci, A., Giorgetta, C., Van't Wout, M., Bonini, N., & Sanfey, A. G. (2012). Reappraising the ultimatum: An fMRI study of emotion regulation and decision making. *Cerebral Cortex*, doi: 10.1093/cercor/bhs028 (in press)
- Hamann, S. (2011). Affective neuroscience: Amygdala's role in experiencing fear. *Current Biology*, 21(2), R75–R77.
- Halali, E., Bereby-Meyer, Y., & Meiran, N. (2011). When rationality and fairness conflict: The role of cognitive-control in the Ultimatum Game. Available at SSRN: <http://ssrn.com/abstract=1868852>
- Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, 324(5927), 646–648.
- Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *Neuroimage*, 30(1), 313–324.
- Hariri, A. R., Mattay, V. S., Tessitore, A., Fera, F., & Weinberger, D. R. (2003). Neocortical modulation of the amygdala response to fearful stimuli. *Biological Psychiatry*, 53(6), 494–501.
- Harlé, K. M., Chang, L. J., van't Wout, M., & Sanfey, A. G. (2012). The neural mechanisms of affect infusion in social economic decision-making: A mediating role of the anterior insula. *Neuroimage*, 61(1), 32–40.
- Harlé, K. M., & Sanfey, A. G. (2007). Incidental sadness biases social economic decisions in the ultimatum game. *Emotion*, 7(4), 876–881.
- Harlé, K. M., & Sanfey, A. G. (2012). Social economic decision-making across the lifespan. *Neuropsychologia*, 50(7), 1416–1424.
- Hewig, J., Kretschmer, N., Trippe, R. H., Hecht, H., Coles, M. G. H., Holroyd, C. B., & Miltner, W. H. R. (2011). Why humans deviate from rational choice. *Psychophysiology*, 48(4), 507–514.
- Johnson, T., Dawes, C. T., Fowler, J. H., McElreath, R., & Smirnov, O. (2009). The role of egalitarian motives in altruistic punishment. *Economics Letters*, 102(3), 192–194.
- Knoch, D., Nitsche, M. A., Fischbacher, U., Eisenegger, C., Pascual-Leone, A., & Fehr, E. (2008). Studying the neurobiology of social interaction with transcranial direct current stimulation—The example of punishing unfairness. *Cerebral Cortex*, 18(9), 1987–1990.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829–832.
- Koenigs, M., & Tranel, D. (2007). Irrational economic decision-making after ventromedial prefrontal damage: Evidence from the ultimatum game. *Journal of Neuroscience*, 27(4), 951–956.
- Lewis, C. M., Baldassarre, A., Comitteri, G., Romani, G. L., & Corbetta, M. (2009). Learning sculpts the spontaneous activity of the resting human brain. *Proceedings of the National Academy of Sciences of the United States of America*, 106(41), 17558–17563.
- Lieberman, M. D. (2011). Why symbolic processing of affect can disrupt negative affect. In A. Todorov, S. Fiske, & D. Prentice (Eds.), *Social neuroscience: Toward understanding the underpinnings of the social mind* (pp. 188–210). Oxford: Oxford University Press.
- Loewenstein, G. F., & O'Donoghue, T. (2004). Animal spirits: Affective and deliberative processes in human behavior. Working Papers 04-14, Cornell University, Center for Analytic Economics.
- MacDonald, A. W., Cohen, J. D., Stenger, V. A., & Carter, C. S. (2000). Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science*, 288(5472), 1835–1838.
- Moll, J., de Oliveira-Souza, R., Eslinger, P. J., Bramati, I. E., Mourão-Miranda, J., Andreiuolo, P. A., & Pessoa, L. (2002). The neural correlates of moral sensitivity: A functional magnetic resonance imaging investigation of basic and moral emotions. *Journal of Neuroscience*, 22(7), 2730–2736.
- Momennejad, I., & Haynes, J. D. (2012). Human anterior prefrontal cortex encodes the 'what' and 'when' of future intentions. *Neuroimage*, 61(1), 139–148.
- Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430.
- Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science*, 314(5805), 1560–1563.
- Paulus, M. P., & Stein, M. B. (2006). An insular view of anxiety. *Biological Psychiatry*, 60(4), 383–387.
- Picton, L., Saunders, B., & Jentsch, I. (2012). "I will fix only my own mistakes": An ERP study investigating error processing in a joint choice-RT task. *Neuropsychologia*, 50(5), 777–785.
- Pillutla, M. M., & Murnighan, J. K. (1996). Unfairness, anger, and spite: Emotional rejections of ultimatum offers. *Organizational Behavior and Human Decision Processes*, 68, 208–224.
- Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, 83, 1281–1302.

- Rilling, J. K., & Sanfey, A. G. (2011). The neuroscience of social decision-making. *Annual Review of Psychology*, 62, 23–48.
- Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, 318(5850), 598–602.
- Sanfey, A. G., Rilling, J. K., Aronson, J. A., Nystrom, L. E., & Cohen, J. D. (2003). The neural basis of economic decision-making in the ultimatum game. *Science*, 300(5626), 1755–1758.
- Schwartz, G. E., & Weinberger, D. A. (1980). Patterns of emotional responses to affective situations: Relations among happiness, sadness, anger, fear, depression, and anxiety. *Motivation and Emotion*, 4(2), 175–191.
- Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, 13(8), 334–340.
- Singer, T., Seymour, B., O'Doherty, J., Kaube, H., Dolan, R. J., & Frith, C. D. (2004). Empathy for pain involves the affective but not sensory components of pain. *Science*, 303(5661), 1157–1162.
- Steinbeis, N., Bernhardt, B. C., & Singer, T. (2012). Impulse control and underlying functions of the left DLPFC mediate age-related and age-independent individual differences in strategic social behavior. *Neuron*, 73(5), 1040–1051.
- Suveg, C., & Zeman, J. (2004). Emotion regulation in children with anxiety disorders. *Journal of Clinical Child and Adolescent Psychology*, 33(4), 750–759.
- Tabibnia, G., Satpute, A. B., & Lieberman, M. D. (2008). The sunny side of fairness: Preference for fairness activates reward circuitry (and disregarding unfairness activates self-control circuitry). *Psychological Science*, 19(4), 339–347.
- Tricomi, E., Rangel, A., Camerer, C. F., & O'Doherty, J. P. (2010). Neural evidence for inequality-averse social preferences. *Nature*, 463(7284), 1089–1091.
- Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, 46, 35–37.
- van't Wout, M., Chang, L. J., & Sanfey, A. G. (2010). The influence of emotion regulation on social interactive decision-making. *Emotion*, 10(6), 815–821.
- van't Wout, M., Kahn, R. S., Sanfey, A. G., & Aleman, A. (2006). Affective state and decision-making in the Ultimatum Game. *Experimental Brain Research*, 169(4), 564–568.
- Wang, J. M., Seidler, R. D., Hall, J. L., & Preston, S. D. (2012). The neural bases of acquisitiveness: Decisions to acquire and discard everyday goods differ across frames, items, and individuals. *Neuropsychologia*, 50(5), 939–948.
- Wright, N. D., Symmonds, M., Fleming, S. M., & Dolan, R. J. (2011). Neural segregation of objective and contextual aspects of fairness. *Journal of Neuroscience*, 31(14), 5244–5252.
- Xiao, E., & Houser, D. (2005). Emotion expression in human punishment behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7398–7401.
- Yamagishi, T., Horita, Y., Takagishi, H., Shinada, M., Tanida, S., & Cook, K. S. (2009). The private rejection of unfair offers and emotional commitment. *Proceedings of the National Academy of Sciences of the United States of America*, 106(28), 11520–11523.
- Zhu, Q., Zhang, J. D., Luo, Y. L., Dilks, D. D., & Liu, J. (2011). Resting-state neural activity across face-selective cortical regions is behaviorally relevant. *Journal of Neuroscience*, 31(28), 10323–10330.

The Fairness Norm in Social Decision-making: Behavioral and Neuroscience Studies

LUO Yi¹; FENG Chunliang¹; GU Ruolei²; WU Tingting¹; LUO Yuejia^{1,3}

(¹ State Key Laboratory of Cognitive Science and Learning, Beijing Normal University, Beijing 100875, China)

(² Institute of Psychology, Chinese Academy of Sciences, Beijing 100101, China)

(³ Chengdu Medical College, Chengdu 610083, China)

Abstract: Widely regarded as a crucial topic in human society, fairness plays an important role in social decision-making. Economists and psychologists focus on discovering the motivational and cognitive mechanism underlying the fairness norm. Based on recent researches on fairness norm, three core models related to the fairness norms, namely inequity aversion, reciprocity, and emotion models, were summarized. Behavioral and neuroscience studies that are for or against these models were reviewed and compared. Specifically, the functions of those brain regions (the insula and the DLPFC) that are closely linked to social decision-making were discussed. Finally, the authors provided their opinions about future research.

Key words: fairness norm; social decision making; emotion; insula; dorsal lateral prefrontal cortex (DLPFC)