

好人也会做坏事：有限道德视角下的不道德行为^{*}

王 芃 王忠军 李松锴

(青少年网络心理与行为教育部重点实验室;

华中师范大学心理学院暨湖北省人的发展与心理健康重点实验室, 武汉 430079)

摘 要 有限道德是指个体在道德判断中固有的局限性, 表现在人们外在行为与内在道德观的悖离, 它导致无意识的道德偏差。有限道德弥补了理性道德模型在不道德行为解释力上的缺陷。该视角下的不道德行为主要包括: 内隐偏见、内群体互惠、沽名钓誉、利益冲突和漠视未来。关于有限道德存有 4 种理论解释: 自利性动机、双重自我、道德褪色和框架效应。未来有关道德决策和不道德行为的研究, 需要延伸到有限理性与无意识层面。

关键词 不道德行为; 理性道德模型; 有限道德; 内群体互惠; 双重自我

分类号 B849: C91

1 引言

不道德行为(unethical behavior)是指违背人们所广泛接受的社会道德规范的行为(Kaptein, 2008; Kish-Gephart, Harrison, & Treviño, 2010)。近年来, 以网络热词为代表的各种商业丑闻、官员腐败、政商合谋、暗箱操作、产品伤害事件屡见媒体, 如“地沟油”、“三氯氰胺”、“表哥”、“房叔”、“小悦悦”等, 引发公众对社会道德和公民素质的激烈争议与拷问。在组织中, 不道德行为逐渐呈现出复杂化、多样化、高科技化和团队化的特点, 其高发性和普遍性不仅给组织造成巨大的经济损失, 也严重破坏企业信誉和品牌形象。

即便存在道德与法律的双重约束, 为什么个体、群体、组织屡屡做出违背普遍伦理的行为? 如何减少甚至消除此类行为? 这些皆属亟需解决的问题。研究不道德行为的原因机制和应对措施, 渐已成为不同学科如心理学、社会学、管理学共同关注的热点。人们把不道德行为归因于道德失范、监管不力、法制缺陷等, 因此在完善制度、加强监管的同时, 道德建设与约束也是减少不道德行为极其重

要的举措。比如高校加大了对学生的职业道德教育, 企业制定了相关措施预防公司员工的道德败坏行为, 行业协会也加强了从业者的自律与道德训练。

然而, 这些措施主要针对显性的不道德行为, 现实中一些不道德行为是人们在无意识中做出的, 很多品行端正的人, 也会不知不觉地做出坏事(Bazerman, Chugh, & Banaji, 2005; Bazerman, Loewenstein, & Moore, 2002; Bersoff, 1999)。例如在缺乏审计独立性时, 审计工作的客观性会受到影响, 审计员会无意识地做出有偏的审计决策, 美国安然公司破产事件即为经典案例(Bazerman et al., 2002; Moore, Tetlock, Tanlu, & Bazerman, 2006); 中国社会盛行的“关系”(guanxi)行为, 因其滋生腐败和损害公平而被视作非道德(Chen, C. C., & Chen, X. P., 2009; Fan, 2002), 但当人们自己处于关系圈内时似乎丝毫未觉“关系”行为有何不妥。显然, 这些是更深层的隐性道德问题, 规范的理性道德模型是难以有效解释的, 通过理性的道德约束也很难控制此类行为。本文引入有限道德的视角, 以解释人们无意识的非道德行为。未来对不道德行为的研究, 需要关注有限理性和无意识层面。

2 不道德行为研究的理论转向

2.1 理性道德模型

理性道德模型假设个体在道德决策过程中是

收稿日期: 2013-01-08

^{*} 国家自然科学基金项目(71232001)、教育部人文社科青年基金项目(10YJC630267)资助。

通讯作者: 王忠军, E-mail: wangzj@mail.ccnu.edu.cn

绝对理性的, 不道德行为是个体在诸多因素的作用下有意而为。理性道德模型是一个广泛的理论框架, 在解释不道德行为的发生机制上, 存在多种理论取向(Kish-Gephart et al., 2010; Maheshwari & Ganesh, 2004)。

第一, 烂苹果理论(bad apples)。此理论取向强调个体因素是影响不道德行为的主要变量。皮亚杰的儿童道德发展阶段理论、柯尔伯格的道德认知发展理论皆认为, 个体的道德发展水平和阶段存在差异。在面临道德困境时, 高道德发展水平的人比低道德发展水平的人, 更可能做出道德的决策。个体在不断社会化过程中所形成的内在道德标准和道德信念, 很大程度上决定了个体的外在道德行为。除道德认知水平外, 其他的个体因素也与不道德行为相关。例如外控型人格比内控型人格更可能导致不道德行为(Forte, 2005); 高马基雅维利主义者比低马基雅维利主义者产生更多的不道德行为(Hegarty & Sims, 1979)。

第二, 坏染缸理论(bad barrels)。该理论取向重视环境因素在引发不道德行为中的作用。Milgram 服从实验、Zimbardo 监狱模拟实验说明, 外界情境可以致使个体做出不道德行为(电击他人和虐待犯人)。环境线索对于不道德行为的催化剂作用在组织中也有所显现。例如道德氛围(ethical climate)反映的是员工对于组织中作为期望标准的某种形式的道德推理和行为的共同知觉(Martin & Cullen, 2006), 与不道德行为显著相关(Kish-Gephart et al., 2010), 处于工具性的道德氛围中的员工较之于处于关爱、制度化、规则化和自主的道德氛围中的员工, 更可能产生不道德行为(Wimbush & Sheperd, 1994); 组织文化(organizational culture)与员工的不道德行为具有相关(Kish-Gephart et al., 2010), 良好正向的组织文化可以减少组织内部不道德行为的发生。

第三, 个人—情境交互模型。Treviño (1986)提出个人 - 情境交互模型, 认为道德决策是建立在道德认知的基础之上, 同时受到个体因素和情境因素的交互作用的结果。当个体面临道德困境时, 决定其是否会产生不道德行为的原因, 除了个体道德认知发展水平以及其他个体因素(如自我强度、场依存性、控制点)之外, 还受到情境因素(如工作背景、组织文化、工作性质)的影响。个人 - 情境交互模型融合了“烂苹果”和“坏染缸”理

论, 是一个较为系统的解释模型。

第四, 问题权变理论。该理论的主要代表是 Jones (1991)所提出的问题权变模型, 它把道德问题本身作为影响道德决策和行为的权变因素。道德决策随具体道德问题而变, 问题的道德强度(moral intensity)不同, 则个体的道德判断亦不同。其中, 道德强度意指道德问题的紧迫程度, 包括 6 个维度: 结果大小(magnitude of consequence), 社会一致性(social consensus), 效应可能性(probability of effect), 时间即时性(temporal immediacy), 亲密性(proximity)和效应集中性(concentration of effect)。通过引入权变思想, 不道德行为的解释模型趋于完善。

2.2 有限道德模型

理性道德模型在解释有意的不道德行为上已较为完善, 然而现实中还有许多现象是其难以有效加以解释的, 如“好心办错事”、“事后后悔”、“无心之过”、“恬不知耻”、“不知者无罪”等。有研究者指出, 很多不道德行为和决策其实是缺乏理性和深思熟虑的, 偏向直觉化、自动化和情绪化(谢熹瑶, 罗跃嘉, 2009; Haidt, 2001; Reynolds, 2006; Sonenshein, 2007)。因此, 研究者开始把视角从理性转向有限理性。

Simon (1956)认为, 人的理性介于完全理性和非理性之间, 受到情境与人们计算能量的制约, 并提出有限理性(bounded rationality)的概念。随后 Kahneman 和 Tversky 继承了 Simon 的观点, 从行为决策的角度提出预期理论(prospect theory), 认为决策者的认知能力和可利用的认知资源是有限的, 在识别和发现问题时, 易受知觉偏差的影响, 从而在运用信息的过程中, 会以系统的方式偏离最佳决策和理性(Kahneman & Tversky, 1979)。根据有限理性和预期理论, 在面临道德冲突和决策时, 个体的理性同样会以系统的方式受到限制, 产生“道德盲点”, 表现出无意识的道德判断与决策上的偏差。

在充分吸收 Simon 和 Kahneman 的理论思想的基础上, Chugh, Bazerman 和 Banaji (2005)系统地提出有限道德的概念, 认为有限道德是有限理性在道德判断与决策领域的迁移, 并将有限道德(bounded ethicality)定义为一种系统的和可预测的心理过程: 它导致人们做出与自身的道德准则相违的不道德行为(Banaji, Bazerman, & Chugh,

2003; Banaji & Bhaskar, 2000)。Chugh 等人(2005)认为在判断和决策过程中, 意识扮演的可能只是一个次要角色, 大多数思维、感觉和动机都是处于无意识模式下。因此, 很多不道德行为实际上是一种“无意的不道德”(unintended unethicity)(Tenbrunsel & Smith-Crowe, 2008)。根据有限道德的观点, 个体外显行为与内在道德信念之间存在断层, 在无意识情境下, 道德认知退变为次要因素, 个体实际表现出来的行为并非他所想的那样道德(Tenbrunsel, Diekmann, Wade-Benzoni, & Bazerman, 2010)。国内学者茅宁和陈银飞(2012)扩展了有限道德的内涵, 认为有限道德是指伦理决策的有限性, 不局限于预测或评价自身的行为上, 还涉及到对他人行为进行伦理判断和评价, 并指出变化视盲、结果偏差和确定效应是有限道德的3种表现(陈银飞, 茅宁, 2011)。目前理性道德框架下的不道德行为研究已较为丰富, 未来研究需要转向有限道德。

3 有限道德视角下的不道德行为表现

有限道德视角下不道德行为的基本特点在于行为主体的无意识性, 个体在从事这些行为的过程中并未觉知其道德性, 但实际上该行为既损害了他人利益也背离了自身道德标准。无意识的不道德行为多种多样, 据国内外学者的研究, 主要有: 内隐偏见、内群体互惠、沽名钓誉、利益冲突和漠视未来(陈银飞, 茅宁, 2009; Banaji et al., 2003; Bazerman, 2011; Bazerman & Banaji, 2004; Tenbrunsel et al., 2010)。

3.1 内隐偏见(implicit discrimination)

个体在决策中通常会存在性别、年龄、种族和文化等偏见, 但个体无法清醒地意识到这些偏见的存在, 即内隐偏见(Chugh et al., 2005)。具体而言, 内隐偏见指对特定群体的无意识偏见, 源自内隐态度, 与外显态度有着明显的区别。外显态度是通过深思熟虑的认知而发展的信念和看法, 是信息加工的有意控制过程, 而内隐态度则是对经验自动反应的心理表征, 无须意识努力。内隐态度常常以细微的方式影响人们的判断、决策和行为。

由于内隐偏见是对特定社会群体和消极评价间的自动认知联系(Fazio, Jackson, & Dunton, 1995), 那么持有内隐偏见的人在受到环境中刺激线索的驱动之后, 会无意识地对相应的社会群体

产生消极的评价和认知, 并在行为中反映出来。例如 Green 等(2007)对医生群体内隐偏见的研究发现, 无意识的种族偏见影响医生对于黑人患者的医疗决策; McConnell 和 Leibold (2001)发现内隐态度能很好地预测人们对待不同群体的非言语行为, 相比对白人主试, 对黑人内隐偏见高的被试显示出了对黑人主试更消极的非言语行为; 内隐联想测验(IAT)的结果也发现某白人男性对于白人/好与黑人/坏的反应时显著短于对于白人/坏与黑人/好的反应时, 尽管他却一再宣称自己是无偏的, 但是在实际工作中却给了白人更多的机会(Bazerman & Moore, 2008)。不同文化环境和被试群体的研究结果证实了内隐偏见的普遍性。虽然表面上我们存在着好意, 但是实际上大多数人都是种族主义者、性别歧视者和老年歧视者(Fiske, 2004)。

3.2 内群体互惠(ingroup favoritism)

社会认同理论认为, 个体通过社会分类, 对自己所属的群体产生认同, 过分热衷于自己的群体, 往往将有利的资源分配给内群体人员(张莹瑞, 佐斌, 2006)。当人们对他们所处的内群体有强烈的认同, 且他们的自尊与群体的价值相联系时, 他们就会帮助自己的群体成员, 有时候甚至会贬低其他的群体(Dasgupta, 2004)。内群体互惠是指由于社会认同的作用, 人们会无意识倾向于帮助内群体人员。划分内外群体的标准通常是人口统计学因素, 人口统计学(如种族、性别、地域等)与己相似的人则被自动归为内群体(Verkuyten, 2007; Mohr & Larsen 1998); 也有研究者直接按照某种标准把被试划分到不同的组别中, 相同组别内的被试属于内群体人员(Tajfel, Billig, Bundy, & Flament, 1971)。事实上, 当我们认为理所应当去帮助自己的朋友、老乡或者亲戚时, 其实已经构成对那些外群体成员的歧视, 导致资源分配产生实质性的不公, 因此从群际层面考虑内群体互惠本质上也是非道德的。它与内隐偏见的不同之处在于, 前者强调偏爱内群体, 后者则侧重于歧视外群体。

关于内群体互惠的研究, 虽然发现在一些特殊情况下可能会出现偏好逆转和害群之马效应(black sheep effect)(Khan & Lambert, 1998), 但是已经证实内群体偏好是普遍存在的(Verkuyten, 2007; Nakashima, Isobe, & Ura, 2008)。例如, Verkuyten (2007)对不同种族的儿童的研究表明,

被试对内群体的评估比对外群体的评估更加积极; Mohr 和 Larsen (1998)对澳大利亚足球比赛中裁判的内群体互惠的研究发现, 本地球队比外地球队得到了更多的任意球判罚; 在组织中, 内群体互惠效应依然存在, 当面临同等合格的两个工作候选人时, 被试更倾向于选择内群体成员(Lewis & Sherman, 2003)。人们的内群体偏好程度越强, 则为内群体分配更多的资源(Oishi & Yoshida, 2002)。当基于个人无意识主观偏好而非客观标准的决策倾向导致资源分配不公时, 内群体互惠已经成为不道德行为。

3.3 沽名钓誉(overclaiming credit)

两个或两个以上的合作者以百分比来评估自己对于其合作成果的贡献, 百分比之和往往超过百分之百。沽名钓誉即指合作双方倾向于高估自己对集体成功的贡献, 同时贬低他人的努力成分(Banaji et al., 2003)。研究者认为, 自我注意聚焦和自我记忆提取的简易性, 造成了人们对于自身贡献的积极幻觉, 即人们倾向于关注自己的贡献而忽视他人的贡献, 同时人们也更容易回忆自身而非他人的贡献(Caruso, Epley, & Bazerman, 2006a)。沽名钓誉的另一根源在于自利性归因偏差, 这种归因偏差使得人们以偏颇的态度来解释事件的成因, 对于集体的成功, 倾向于认为自己有更大的功劳, 而对于集体的失败, 则倾向于认为自己承担很小的责任(Miller & Ross, 1975)。

沽名钓誉现象在一些研究领域已经得到证实(Brawley, 1984; Forsyth & Schlenker, 1977; Ross & Sicoly, 1979)。一项经典的研究是 Ross 和 Sicoly (1979)关于夫妻各自评估自己和对方承担家庭事务中的贡献大小的研究, 研究发现夫妻双方高估了各自的贡献, 认为自己承担了更多的家务。由于沽名钓誉, 人们难以客观评估合作者的贡献, 这可能影响后续合作行为。组织中沽名钓誉现象主要表现在绩效评估中, 组员对自己和他人的绩效判断可能会存在差异, 认为自己的绩效优于其他组员, 很少有人会认为自己的绩效排序在后50%的行列, 这在一定程度上会降低团体凝聚力。由于人们时常处于“当局者迷”的状态中, 易于居功自傲, 自认合乎事实, 实际上却夸大了个人努力, 贬低了他人成就。积极地思考其他组员的观点, 采取他人的视角, 可能是减少这些无意识的沽名钓誉现象的一个有效手段(Caruso et al.,

2006a; Caruso, Epley, & Bazerman, 2006b)。

3.4 利益冲突(conflict of interest)

医生道德沦丧的原因大多在于医药不分家, 以药补医, 经纪人在股市萧条时会仍然建议客户大量持有股票, 审判独立性要求法官回避审判与法官有利益关联的人。现实生活中利益关联现象普遍存在并且影响到人们的行为决策。尽管人们试图保持客观, 当存在个人利益卷入时, 人们的决策和判断便会受到影响, 常常未能意识到自身所面临的利益冲突, 从而导致不道德行为(Moore et al., 2006; Thagard, 2007)。简言之, 欲望与潜在利益瓜葛会影响人们解释信息的方式, 其行为决策的结果往往导致相关利益者蒙受损失。

有关利益冲突的实证研究以审计师判断与决策行为研究为代表。Bazerman 等(2002)进行的公司价值评估实验发现, 扮演不同角色的审计人员(买者和卖者), 根据相同信息对公司客观价值的评估是不同的, 扮演卖方审计人员的估计值要显著高于扮演买方审计人员的估计值。Moore, Tanlu 和 Bazerman (2010)的另一项研究发现, 假如审计师所审计的公司同时也是雇佣他的公司, 而非第三方, 那么该审计师更可能认为该公司的潜在问题账务符合 GAAP (公认会计原则)的要求。当缺乏审计独立性时, 审计人员的判断极有可能受到个人利益冲突的影响。虽然利益冲突不太可能导致有意的不道德行为, 但它有可能不知不觉地扭曲我们的判断。诸多文献都已揭示审计人员具有审计偏差(Bazerman, Moore, Tetlock, & Tanlu, 2006; Moore et al., 2006, 2010), 即便公开和严格的审计规则都无法消除审计中的自利性偏差, 只有消除产生自利性偏差的刺激才能够有效的减少审计人员的无意识偏差, 即减少审计师与客户之间的利益关联。例如美国的 Sarbanes-oxley 法案就禁止审计公司提供给他们所审计的企业相关的咨询服务。Bazerman 等(2002)的建议是: 要想做到真正的审计独立, 消除审计人员的利益冲突, 首先必须确保咨询和税务服务完全独立, 其次, 审计人员必须拥有固定的合约, 在合约期内企业不得终止审计师的服务, 所有的账务和合约细节都必须具体化并且不可更改。

3.5 漠视未来(discounting the future)

我们在预期将来是否会从事保护环境的行为时, 常常信誓旦旦地表明自己会践行环境友好主

义,事实上我们却不是合格的环保者。譬如虽然在一年内就能收回额外成本,但是很多人却不愿意购买高价节能电器。当人们偏好于当前自身的微小的即时利益而忽视将来他人更大的利益时,便产生了代际折扣(intergenerational discounting)(Wade-Benzoni & Tost, 2009)。人们倾向于对属于他人的价值而非自己的价值产生折扣(Loewenstein, Thompson, & Bazerman, 1989),同时,人们对延迟的价值而非即时的价值产生折扣(Loewenstein, 1992),这两种类型的折扣效应引发了代际折扣。漠视未来的实质是处于代际困境的人们在决策的过程中产生代际折扣,漠视下一代人的利益。Wade-Benzoni (2002)认为代际间存在着利益的冲突,当代人的决策和行为会使自身受益却牺牲下一代的利益,通常表现在追求经济利益而造成对生态环境的破坏影响到下一代的健康与生存上。显然,基于代际层面的考虑,漠视未来从属不道德行为范畴。

此外,建构水平加工理论从心理表征水平的角度阐释了时间距离与选择偏好的关系,该理论认为人们对事件建构分为两个水平:对首要特征的高水平建构,它是概括性的和去情境化的,对次要特征的低水平建构,它是具体的和情境性的。时间距离的变化可以改变人们的建构水平,对未来事物是高水平建构,对眼前事物是低水平建构,从而导致决策结果就截然不同(Liberman & Trope, 1998)。由于建构水平的不同,人们基于当前时间点下对未来行为的预测易于产生偏差。例如, Mitchell, Thompson, Peterson 和 Cronk (1997)研究表明,人们对事件的预期比他们在事件进行时的实际体验要更加积极;人们预测将来的行动时,倾向于认为自己比别人表现得更道德(Epley, & Dunning, 2006; Diekmann, 2008)。因此,人们在预测将来行动时倾向于认为自己是个道德的环保主义者,但在实际情形中依然看重即时利益,违背个人准则。

以上列举的各种无意识不道德行为,推翻了理性道德假设。以后对于不道德行为的研究需要拓展视野,从意识层面转向无意识层面,从无限道德假设转向有限道德。

4 有限道德的理论解释

4.1 自利性动机

人们存在维护积极自尊的自利性动机,比如

人们通常把自己看作是道德的、有能力的和有价值的(Chugh et al., 2005),相信自己比其他人更可能表现出无私、友好和慷慨的行为(Epley & Dunning, 2000),并且固执地认为自己在各种测试和任务中能够完成得比其他人更好(Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995),相比于对他人,我们对于自己的道德行为给予更多的奖赏,而对于自己的不道德行为则承担更小的责任(Messick & Bazerman, 1996)。这种主观倾向性和积极错觉,使得个体出于维持积极的自我概念的目的,而不能有效地监控自己的行为。

通常自利性动机催使人们在积极人格特质上的评估产生自我服务偏差,对自我的评估要优于对他人的评估。在自我道德性评估上亦是如此,人们倾向于把自身看得比一般人更加道德(Tenbrunsel et al., 2010),产生中等偏上效应(above-average effect)。人们总是认为自己是道德的和无偏的,并且想象自己是个合格的决策者,能够客观公正且有理性地做出决策,但是事实是我们的无意识内隐偏见总是与意识层面的外显信念不一致,甚至最善意的人也会不知不觉地允许无意识想法和感觉影响似乎客观的决定(Banaji et al., 2003)。对自身的道德性过于自信,在进行道德决策时未能意识到自身局限性,易产生有偏的道德决策(Chugh et al., 2005; Tenbrunsel et al., 2010)。

4.2 双重自我

Strotz (1955-1956)首次提出双重自我问题(two-selves problem),即人们具有不同的自我,这两种自我在当前和将来两个不同的时期,所起作用的权重是不一样的(Schelling, 1984)。Bazerman, Tenbrunsel 和 Wade-Benzoni (1998)认为在内心冲突时人脑中存在两种自我:“想要的我”和“应该的我”。“想要的我”具有情绪性、情感性、冲动性,被视为“热向的”(hot headed);“应该的我”则更多的是理智的、认知的、深思的,被看作“冷向的”(cool headed)。“应该的我”包含了道德意图在其中,要求人们根据道德准则行事,“想要的我”则包含了自利性,相对忽视了道德性。两种自我之间存在碰撞与斗争,人们的内心冲突来自于“想要的我”与“应该的我”之间的张力。

Tenbrunsel 等人(2010)通过“想要的我/应该的我”框架来解释人们的知行不一,认为这两种

自我在过去、现在和将来不同时间段的主导作用是不同的。在预测和回忆行为的时候,“应该的我”往往占主导地位;在实际行动的时候,“想要的我”往往占主导地位。因此,人们会错误地预测将来自己在道德困境中的行为表现,过高的估计自己会表现出社会赞许行为(Epley & Dunning, 2000),错误地回忆过去已发生的行为,倾向于认为自己比实际上更道德(Mitchell et al., 1997)。这两种自我在不同阶段的主导作用使得人们的实际行动和自身信念产生偏差,导致有限道德的产生(陈银飞,茅宁,2009; Tenbrunsel et al., 2010)。

Woodzicka 和 LaFrance (2001)在工作面试情境下检验了被试在想象和真实情境下对性骚扰的反应,结果发现两组的反应存在巨大差异,想象组的被试都报告说自己会采取措施(例如退出面试、上报主管等),而真实组的被试没有采取任何措施。对此, Tenbrunsel 等人(2010)的解释是,想象组被试在面临决策时是“应该的我”处于主导地位,因此会按照社会期望来反应,而真实组则是“想要的我”处于主导地位,获得工作是被试的主要愿望,因此会选择沉默。

4.3 道德褪色

道德褪色(ethical fading)是 Tenbrunsel 和 Messick (2004)提出的概念,它是一种心理过程,意指在实际的行动过程中,人们没有意识到他所决策的情境具有道德意义,因此就没有把道德标准纳入决策标准之中(Tenbrunsel et al., 2010)。Tenbrunsel 和 Messick (2004)认为,道德褪色是通过自我欺骗(self-deception)过程得以实现的。自我欺骗包括 4 种方式:委婉言语(language euphemisms)、决策滑坡(slippery slope of decision making)、因果知觉谬误(errors in perceptual causation)和自我表征的限制(constrained representation of self)。道德决策通常涉及自利和道德准则的权衡,自我欺骗使得个体避免或者掩饰决策中的道德成分,个体实际上行为自利但仍然认为自己保持着道德性的完整。

Rest (1986)提出的模型把道德行为的产生分为 4 个阶段:第一,意识到道德问题,即个体认识到决策情景存在着道德两难的阶段;第二,做出道德判断,即个体通过多种策略来决定某个行为从道德上来说是否正确还是错误的阶段;第三,确定道德意图,即个体决定以一种道德还是不道德

的方式来行动的阶段;第四,实施道德行为,即个体实施道德或不道德行为的阶段。在该四阶段模型中,道德意识指的是决策者对当前问题情境是否存在道德意义的识别程度,它是道德行为的先行因素。按照此模型,决策者在第一阶段即道德意识阶段,未能意识到所决策的内容包含道德意义,产生道德褪色,最终导致了不道德行为。

4.4 框架效应

框架效应概念来自于行为决策领域。Tversky 和 Kahneman (1981)用“框架效应”来描述决策问题中因情景表述的改变而导致选择偏好改变的现象,并采用经典的“亚洲疾病”问题研究范式证明了框架效应的存在,被试由于相同情况的呈现方式不同,产生了截然不同的决策。在获利框架下,被试倾向于风险规避,在损失框架下,被试产生了偏好逆转,倾向于冒险,即风险寻求。这说明在问题界定与表征过程中,框架对于个体判断和决策的影响。

在前人的理论上, Kern 和 Chugh (2009)的研究发现,道德决策也易受到自动框架的影响,当进行道德决策时,如果个体把潜在的结果知觉为损失的话,他会由于避免损失而比知觉为收益的个体进行更多的不道德行为。另一方面, Tenbrunsel 和 Messick (1999)的研究发现,被试在社会困境中的合作行为,依赖于被试对决策类型的框定,不同的框定类型(经济框架和道德框架)会产生不同的行为结果,把决策内容知觉为道德问题的被试比把决策内容知觉为经济问题的被试,在囚徒困境中更可能表现出合作行为和道德行为; Tenbrunsel 和 Smith-Crowe (2008)也认为,在不同的框架类型下个体的加工是不同的,当人们运用道德框架而非其他框架(经济框架、法律框架)框定决策问题时,他们就更能意识到问题的道德性,反之则无道德意识,更可能产生有限道德和不道德行为。概括而言,个体在面临决策时会自发的运用内在的框架来界定问题的性质,不同的框架会产生不同的决策。因此,由于框架效应的存在,人们会无意识地在经过自己的内在框定之后再行决策,如果采取的是一种非道德框架,那么他们很可能会表现出不道德行为。

5 总结与展望

第一,已有的道德理论模型并不能解释现实

中所有的不道德行为,故而研究视角转向有限道德。研究视角的转变既完善了道德理论,又指出了未来的干预途径和方向。以往道德理论家重视道德训练和道德教育对减少人们不道德行为的作用,但是传统道德训练的效果是有限的(Banaji et al., 2003)。当我们认为自己比实际上更道德时,我们就不太可能去寻找提高我们道德行为的方式,因此 Tenbrunsel 等(2010)认为人们应当首先意识到有限道德的存在,并且承认我们对无意识偏见的易受影响性,才有可能纠正自身无意识的不道德行为。鉴于此,我们在考虑道德应对措施时,就不能够仅仅依赖道德训练或者行为法则的功效,同时要让那些自认为不会从事不道德行为的人认识到理性的局限。

第二,社会发展引发不道德行为呈现多样化趋势,除了前文介绍到的5种无意识的不道德行为之外,是否存在更多类似的不道德行为可以通过有限道德来解释?不同的文化背景下的不道德行为是否存在差异?例如,中国儒家文化孕育的“关系”行为在内涵上与内群体互惠有诸多相似,其盛行的原因除了法制的完善之外,关系行为的双方相互处于内群体之中,产生内群体互惠,人们在“走关系”的过程中可能没有意识到自己的行为实际上却损害了外群体的利益。有限道德视角为探讨中国的“关系”行为的机制提供了某种线索。未来研究可以关注无意识不道德行为的跨文化差异和本土化特征。

第三,有限道德的4种理论解释,其共同点都旨在说明无意识不道德行为产生的心理过程,但各有侧重。自利性动机强调个体维持积极自尊的动机因素,双重自我则以二分的方式强调不同自我的主导作用,道德褪色指出个体道德意识的缺乏,框架效应则强调个体对问题情境的知觉方式。纵观全文可以推之,沽名钓誉与利益冲突的原因在于自利性动机,漠视未来是由于双重自我的效应,内隐偏见与内群体互惠则与道德褪色相关。不同类型的无意识不道德行为对应着不同的心理加工过程,有限道德正是提供了这样一个框架综合了各种类型的无意识不道德行为。无意识的不道德行为有别于有意的不道德行为,其产生机制更多依赖的是阈下自动加工过程,现阶段对于有限道德成因的探讨多是理论假设,未来需要更多可操作的实证研究来检验这些理

论的解释效力。

第四,有限道德的验证性研究目前仍然缺少范式。根据有限道德的定义:当事人的道德认知与外显行为矛盾。因此验证有限道德的存在除了测量个体的行为之外,还需要测量其道德信念与道德意识,通过发现二者的差异以证明有限道德的存在。但是有可能存在这样的状况:虽然个体在理智状态的道德认知是高级的,实际行为也是非道德的,但是此时我们如果直接推断该个体产生有限道德仍然存在风险,因为该个体在进行道德决策时有可能意识到了行为的不道德性,但是由于利益关系仍然“有意而为之”。如何确定不道德行为的无意识性和剥离意识性的影响应当是未来研究需要解决的问题。另外,如果采取继时性测量方式先后测量道德认知和道德行为,则易造成道德启动,影响被试后续的作答和行为表现。未来研究可以采用纵向研究或者多源数据收集的方式减少被试道德意识与道德行为之间的相互作用,以验证有限道德的存在。

第五,由于人们普遍具有社会赞许倾向,所以道德问题研究一直存在方法论上的缺陷,传统的问卷和情景模拟并不足以作为衡量个体道德水平的精确指标,未来研究需要多元化测量方式,例如内隐联想测验(IAT)。此外,一些研究者已经在使用脑成像技术来观察人们在道德决策时大脑区域的激活程度。在执行道德判断任务时,通过fMRI发现腹内侧前额叶、杏仁核、内侧前额叶、扣带后回等相关脑区具有不同程度的激活(Harenski & Hamann, 2006; Moll, Eslinger, & de Oliveira-Souza, 2001)。随着社会认知神经科学的发展,更多生理反馈技术被应用于社会问题的探究,通过认知科学方法和社会研究方法的结合来揭示道德判断的神经机制。未来研究可以利用更加高级的技术以客观的生理指标来反映无意识道德决策的过程和特点。

参考文献

- 陈银飞, 茅宁. (2009). 从有限理性到有限道德——论伦理决策的有限性. *江苏大学学报(社会科学版)*, 11(6), 89-92.
- 陈银飞, 茅宁. (2011). 有限道德、伦理判断与供应商伦理管理: 一个综合研究框架. *华东经济管理*, 25(9), 113-117.
- 陈银飞, 茅宁. (2012). 有限道德、确定效应与供应商伦理

- 管理的实验研究. *商业经济与管理*, (1), 65–72.
- 谢熹瑶, 罗跃嘉. (2009). 道德判断中的情绪因素——从认知神经科学的角度进行探讨. *心理科学进展*, 17(6), 1250–1256.
- 张莹瑞, 佐斌. (2006). 社会认同理论及其发展. *心理科学进展*, 14(3), 475–480.
- Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J., & Vredenburg, D. S. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology*, 68, 804–825.
- Banaji, M. R., Bazerman, M. H., & Chugh, D. (2003). How (un) ethical are you? *Harvard Business Review*, 81(12), 56–64, 125.
- Banaji, M. R., & Bhaskar, R. (2000). Implicit stereotypes and memory: The bounded rationality of social beliefs. In D. L. Schacter, & E. Scarry (Eds.), *Memory, brain, and belief* (pp. 139–175). Cambridge, MA: Harvard University Press.
- Bazerman, M. H. (2011). Bounded ethicality in negotiations. *Negotiation and Conflict Management Research*, 4(1), 8–11.
- Bazerman, M. H., & Banaji, M. R. (2004). The social psychology of ordinary ethical failures. *Social Justice Research*, 17(2), 111–115.
- Bazerman, M. H., Chugh, D., & Banaji, M. R. (2005). When good people (seem to) negotiate in bad faith. *Negotiation*, 8, 1–4.
- Bazerman, M. H., Loewenstein, G., & Moore, D. A. (2002). Why good accountants do bad audits. *Harvard Business Review*, 80(11), 96–102, 134.
- Bazerman, M. H., & Moore, D. A. (2008). *Judgment in managerial decision making*. New York: John Wiley.
- Bazerman, M. H., Moore, D. A., Tetlock, P. E., & Tanlu, L. (2006). Reports of solving the conflicts of interest in auditing are highly exaggerated. *Academy of Management Review*, 31(1), 43–49.
- Bazerman, M. H., Tenbrunsel, A. E., & Wade-Benzoni, K. (1998). Negotiating with yourself and losing: Making decisions with competing internal preferences. *Academy of Management Review*, 23(2), 225–241.
- Bersoff, D. M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin*, 25(1), 28–39.
- Brawley, L. R. (1984). Unintentional egocentric biases in attributions. *Journal of Sport Psychology*, 6, 264–278.
- Caruso, E. M., Epley, N., & Bazerman, M. H. (2006a). The costs and benefits of undoing egocentric responsibility assessments in groups. *Journal of Personality and Social Psychology*, 91(5), 857–871.
- Caruso, E. M., Epley, N., & Bazerman, M. H. (2006b). The good, the bad, and the ugly of perspective taking in groups. In A. E. Tenbrunsel (Ed.), *Ethics in Groups (Research on Managing Groups and Teams)* (Volume 8, pp. 201–224). Emerald Group Publishing Limited.
- Chen, C. C., & Chen, X. P. (2009). Negative externalities of close guanxi within organizations. *Asia Pacific Journal of Management*, 26(1), 37–53.
- Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In D. A. Moore, D. M. Cain, G. F. Loewenstein, & M. H. Bazerman (Eds.), *Conflicts of interest: Problems and solutions from law, medicine and organizational settings* (pp. 74–95). London: Cambridge University Press.
- Dasgupta, N. (2004). Implicit ingroup favoritism, outgroup favoritism, and their behavioral manifestations. *Social Justice Research*, 17(2), 143–169.
- Diekmann, K. A. (2008). "She did what? There is no way I would do that!" The potential interpersonal harm caused by mispredicting one's behavior. *Journal of Business Ethics*, 80(1), 5–11.
- Epley, N., & Dunning, D. (2000). Feeling "holier than thou": Are self-serving assessments produced by errors in self-or social prediction? *Journal of Personality and Social Psychology*, 79(6), 861–875.
- Epley, N., & Dunning, D. (2006). The mixed blessings of self-knowledge in behavioral prediction: Enhanced discrimination but exacerbated bias. *Personality and Social Psychology Bulletin*, 32(5), 641–655.
- Fan, Y. (2002). Ganxi's consequences: Personal gains at social cost. *Journal of Business Ethics*, 38, 371–380.
- Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology*, 69(6), 1013–1027.
- Fiske, S. T. (2004). Intent and ordinary bias: Unintended thought and social motivation create casual prejudice. *Social Justice Research*, 17(2), 117–127.
- Forsyth, D. R., & Schlenker, B. R. (1977). Attributional egocentrism following performance of a competitive task. *The Journal of Social Psychology*, 102(2), 215–222.
- Forte, A. (2005). Locus of control and the moral reasoning of managers. *Journal of Business Ethics*, 58, 65–77.
- Green, A. R., Carney, D. R., Pallin, D. J., Ngo, L. H., Raymond, K. L., Iezzoni, L. I., & Banaji, M. R. (2007). Implicit bias among physicians and its prediction of thrombolysis decisions for black and white patients. *Journal of General Internal Medicine*, 22(9), 1231–1238.
- Haidt, J. (2001). The emotional dog and its rational tail: A

- social intuitionist approach to moral judgment. *Psychological Review*, 108(4), 814–834.
- Harenski, C. L., & Hamann, S. (2006). Neural correlates of regulating negative emotions related to moral violations. *NeuroImage*, 30, 313–324.
- Hegarty, W. H., & Sims, H. P. (1979). Organizational philosophy, policies, and objectives related to unethical decision behavior: A laboratory experiment. *Journal of Applied Psychology*, 64(3), 331–338.
- Jones, T. M. (1991). Ethical decision making by individuals in organizations: An issue-contingent model. *Academy of Management Review*, 16(2), 366–395.
- Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2), 263–292.
- Kaptein, M. (2008). Developing a measure of unethical behavior in the workplace: A stakeholder perspective. *Journal of Management*, 34(5), 978–1008.
- Kern, M. C., & Chugh, D. (2009). Bounded ethicality: The perils of loss framing. *Psychological Science*, 20, 378–384.
- Khan, S., & Lambert, A. J. (1998). Ingroup favoritism versus black sheep effects in observations of informal conversations. *Basic and Applied Social Psychology*, 20(4), 263–269.
- Kish-Gephart, J. J., Harrison, D. A., & Treviño, L. K. (2010). Bad apples, bad cases, and bad barrels: Meta-analytic evidence about sources of unethical decisions at work. *Journal of Applied Psychology*, 95(1), 1–31.
- Lewis, A. C., & Sherman, S. J. (2003). Hiring you makes me look bad: Social-identity based reversals of the ingroup favoritism effect. *Organizational Behavior and Human Decision Processes*, 90(2), 262–276.
- Liberman, N., & Trope, Y. (1998). The role of feasibility and desirability considerations in near and distant future decisions: A test of temporal construal theory. *Journal of Personality and Social Psychology*, 75, 5–18.
- Loewenstein, G. (1992). The fall and rise of psychological explanations in the economics of intertemporal choice. In G. Loewenstein, & J. Elster (Eds.), *Choice over time*. New York: Russell Sage.
- Loewenstein, G. F., Thompson, L., & Bazerman, M. H. (1989). Social utility and decision making in interpersonal contexts. *Journal of Personality and Social Psychology*, 57, 426–441.
- Maheshwari, S. K., & Ganesh, M. P. (2004). *Ethics in organizations: The case of tata steel*. Working Papers.
- Martin, K. D., & Cullen, J. B. (2006). Continuities and extensions of ethical climate theory: A meta-analytic review. *Journal of Business Ethics*, 69, 175–194.
- McConnell, A. R., & Leibold, J. M. (2001). Relations among the implicit association test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of Experimental Social Psychology*, 37(5), 435–442.
- Messick, D. M., & Bazerman, M. H. (1996). Ethics for the 21st century: A decision making approach. *Sloan Management Review*, 37, 9–22.
- Miller, D. T., & Ross, M. (1975). Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, 82(2), 213–225.
- Mitchell, T. R., Thompson, L., Peterson, E., & Cronk, R. (1997). Temporal adjustments in the evaluation of events: The "rosy view". *Journal of Experimental Social Psychology*, 33(4), 421–448.
- Mohr, P. B., & Larsen, K. (1998). Ingroup favoritism in umpiring decisions in Australian football. *The Journal of Social Psychology*, 138(4), 495–504.
- Moll, J., Eslinger, P. J., & de Oliveira-Souza, R. (2001). Frontopolar and anterior temporal cortex activation in a moral judgment task: preliminary functional MRI results in normal subjects. *Arq Neuropsiquiatr*, 59, 657–664.
- Moore, D. A., Tanlu, L., & Bazerman, M. H. (2010). Conflict of interest and the intrusion of bias. *Judgment and Decision Making*, 5(1), 37–53.
- Moore, D. A., Tetlock, P. E., Tanlu, L., & Bazerman, M. H. (2006). Conflicts of interest and the case of auditor independence: Moral seduction and strategic issue cycling. *The Academy of Management Review*, 31(1), 10–29.
- Nakashima, K., Isobe, C., & Ura, M. (2008). Effect of self-construal and threat to self-esteem on ingroup favoritism: Moderating effect of independent/interdependent self-construal on use of ingroup favoritism for maintaining and enhancing self-evaluation. *Asian Journal of Social Psychology*, 11(4), 286–292.
- Oishi, C., & Yoshida, F. (2002). Black sheep effect and ingroup favoritism in social identity perspective. *The Japanese Journal of Psychology*, 73, 405–411.
- Rest, J. R. (1986). *Moral development: Advances in research and theory*. New York: Praeger.
- Reynolds, S. J. (2006). A neurocognitive model of the ethical decision-making process: Implications for study and practice. *Journal of Applied Psychology*, 91(4), 737–748.
- Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, 37(3), 322–336.
- Schelling, T. C. (1984). *Choice and consequence: Perspectives of an errant economist*. Cambridge, MA: Harvard University Press.
- Simon, H. A. (1956). Rational choice and the structure of the

- environment. *Psychological Review*, 63(2), 129–138.
- Sonenshein, S. (2007). The role of construction, intuition, and justification in responding to ethical issues at work: The sensemaking-intuition model. *The Academy of Management Review*, 32(4), 1022–1040.
- Strotz, R. H. (1955-1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, 23, 165–180.
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology*, 1(2), 149–178.
- Tenbrunsel, A. E., Diekmann, K. A., Wade-Benzoni, K. A., & Bazerman, M. H. (2010). The ethical mirage: A temporal explanation as to why we are not as ethical as we think we are. *Research in Organizational Behavior*, 30, 153–173.
- Tenbrunsel, A. E., & Messick, D. M. (1999). Sanctioning systems, decision frames, and cooperation. *Administrative Science Quarterly*, 44, 684–707.
- Tenbrunsel, A. E., & Messick, D. M. (2004). Ethical fading: The role of self-deception in unethical behavior. *Social Justice Research*, 17(2), 223–236.
- Tenbrunsel, A. E., & Smith-Crowe, K. (2008). Ethical decision making: Where we've been and where we're going. *The Academy of Management Annals*, 2(1), 545–607.
- Thagard, P. (2007). The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy*, 24(4), 367–380.
- Treviño, L. K. (1986). Ethical decision making in organizations: A person-situation interactionist model. *Academy of Management Review*, 11(3), 601–617.
- Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, 211(4481), 453–458.
- Verkuyten, M. (2007). Ethnic in-group favoritism among minority and majority groups: Testing the self-esteem hypothesis among preadolescents. *Journal of Applied Social Psychology*, 37(3), 486–500.
- Wade-Benzoni, K. A. (2002). A golden rule over time: Reciprocity in intergenerational allocation decisions. *Academy of Management Journal*, 45(5), 1011–1028.
- Wade-Benzoni, K. A., & Tost, L. P. (2009). The egoism and altruism of intergenerational behavior. *Personality and Social Psychology Review*, 13(3), 165–193.
- Wimbush, J. C., & Shepard, J. M. (1994). Toward an understanding of ethical climate: Its relationship to ethical behavior and supervisory influence. *Journal of Business Ethics*, 13, 637–647.
- Woodzicka, J. A., & LaFrance, M. (2001). Real versus imagined gender harassment. *Journal of Social Issues*, 57, 15–30.

Good People Do Bad Things: Unethical Behavior from the Perspective of Bounded Ethicality

WANG Peng; WANG Zhongjun; LI Songkai

(Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education;
School of Psychology, Central China Normal University; Key Laboratory of Human Development and
Mental Health of Hubei Province, Wuhan 430079, China)

Abstract: Bounded ethicality refers to a kind of inherent limitation when individuals make ethical decision. It manifests the departure between the external behavior and internal moral value, and leads to unconscious moral bias. Combined with rational ethicality model, bounded ethicality is an alternative explanation of unethical behavior. From the perspective of bounded ethicality, unethical behavior includes implicit bias, in-group favoritism, over-claiming credit, conflicts of interest and discounting the future. And the theoretical explanation arises from self-service motive, two-selves, ethical fading and frame effect. Studies about ethical decision making and unethical behavior in the future need to be extended to bounded rationality.

Key words: unethical behavior; rational ethicality model; bounded ethicality; in-group favoritism; two-selves