

# 自我欺骗：通过欺骗自己更好地欺骗他人

陆慧菁

(香港理工大学应用社会科学系, 香港)

**摘要** 心理学的研究几乎都把自欺看作一种个体特质, 主要探讨自欺对个体自身的作用。而在进化的理论框架中, 自欺是一种人际交往的策略, 目的是为了能够更好地欺骗其他个体。由于直接欺骗有可能被对方识破, 个体把自己欺骗之后就可以“诚实地”向对方传递错误信息而不被察觉。这就把自欺定义为人际间而非个体内的概念。本文系统地阐述人际间自欺的概念和思路, 并为之在心理学的研究中提出一套理论假设和研究方法。在理论上, 由于难以区别通过自欺的欺骗和直接的欺骗, 本文提出一系列假设和条件来证明自欺的存在, 这些条件包括欺骗双方的地位高低、道德高低、和人数多寡。因为自欺是用来应付高欺骗探测压力的, 上述条件可以反映被骗对象的欺骗探测能力的高低。在方法上, 本文提出采用两次回忆的方式来展现自欺, 第一次回忆时欺骗动机存在, 个体在诚实的状态下提供较少真实信息, 第二次回忆时欺骗动机消失, 个体可以提供较多真实信息, 通过比较两次回忆内容的差异来验证自欺的存在。这一操作方法突破了个体间自欺因缺乏操作定义而无法进行心理学研究的屏障, 为人们探讨人际间自欺提供了一条可行途径。

**关键词** 自我欺骗; 欺骗; 进化心理学

**分类号** B849:C91

心理学领域中关于自我欺骗(self-deception)的研究多数集中于自欺如何影响个体的情绪、认知、动机、决策等等(Balcetis, 2008; Erez, Johnson, & Judge, 1995; Hirschfeld, Thomas, & McNatt, 2008; Triandis, 2009), 通常采用问卷(Paulhus, 1991; Sackeim & Gur, 1979)来测量自欺, 并讨论适度自欺的功能和作用, 过度自欺的弊端等。这是从个体内的角度(intrapersonal)来看待自欺, 把自欺当作一种个体特质。而在心理学研究中, 有些心理状态可以是个体的一种长期特质, 同时又是一种可被诱发的暂时状态, 例如情绪低落, 长期的可以是抑郁, 短期的可以是任务失败后的感受。那么, 自欺能否被当作一种短暂的状态或策略来研究呢? 心理学中似乎没有相关的理论和假设。

从生物进化的角度看, 个体欺骗自己, 使得自己接触不到真实信息, 似乎不利于个体适应性的提高, 因为真实信息对个体来说往往是有利的(Wallace, 1973)。那么, 自欺这种看似不适宜的行为, 是如何发展出来并被自然选择保留至今的呢? 进

化生物学家 Robert Trivers 在理察·道金斯的《自私的基因》一书的序言中提出, 自欺的出现是由于个体为了更好地欺骗其他个体(Trivers, 1976), 只要骗人的得益大于自欺的损失, 自欺就可以不被自然选择淘汰掉。这是从个体间的角度(interpersonal)来诠释自欺, 把自欺看作一种个体之间互动的策略。两只猛兽相遇, 尽管实际上两者的战斗力可能不相上下, 自认为强悍(即自欺)的一方竖起毛发虚张声势, 就容易吓跑对方(成功欺骗)。这个源于生物学的思想, 对研究人类自欺, 尤其是作为人际交往策略的自欺, 有很大的启发作用。

然而, 把自欺当作一种欺骗策略来研究, 其困难之一是区分直接欺骗和通过自欺来欺骗。心理学中没有这方面的理论假设。本文根据人际间欺骗的思路, 提出欺骗对象的欺骗探测能力是影响自欺的重要因素, 如果对方的欺骗探测能力高, 个体就需要通过自欺来欺骗, 如果对方的欺骗探测能力低, 个体可以直接欺骗。本文还提出一系列影响探测能力的情境和条件, 包括欺骗双方的地位高低、道德

收稿日期: 2011-09-22

通讯作者: 陆慧菁, E-mail: huijinglu@gmail.com

1265

高低、和人数多寡。有了这些情境,研究者就可以比较自欺和直接欺骗。困难之二是测量无意识,因为自欺者意识不到自己正在进行欺骗,似乎只能通过探测无意识来证明自欺的存在。本文提出通过比较两次回忆内容的差异来显示个体运用了自欺策略,这种办法既绕过了测量无意识的困难,又给人际间的自欺下了一个操作性定义,并可以方便应用到人际间欺骗的情境中。文章分三部分,在从个体间的角度讨论自欺之前,先概述从个体内角度来探讨自欺的研究(第一部分),接着阐述进化视角下的自欺理论,并提出可验证该理论的情境及假设(第二部分),最后提出研究人际间自欺的方法(第三部分)。

## 1 从“个体内”的角度探讨自欺

### 1.1 自欺现象和哲学上的争论

可能有很多现象都会被人们称作自欺,比如盗铃的人以为掩着自己的耳朵别人就听不见他偷铃,吃不到葡萄的狐狸就说葡萄是酸的,没穿衣服的皇帝以为自己穿了一件奇特的新装,被人打了的阿Q认为自己是被儿子打的,小偷的母亲认为自己的儿子是善良的人不是盗贼,患病的人认为自己没病,是医生诊断错误,等等。这些现象通常是自欺的人持有与事实不相符的看法。在心理学研究中,有一些现象也与自欺有关。例如,只得到少量报酬的被试向他人谎称无聊的任务有趣后,对任务的趣味性评价提高(Festinger & Carlsmith, 1959),这个过程借助自欺来完成(Shapiro, 1996);声称对黑人没有敌意的白人在反应时任务中表现出对黑人的回避和负性认知(Greenwald, McGhee, & Schwartz, 1998);人们从过分正面的角度来看待自己的能力和道德水平(Paulhus & John, 1998);个体选择性地接收信息或自利性地诠释信息等有偏差的认知加工过程(Greenwald, 1988);等等。

在哲学上,自欺是否成立一直备受争议。因为自己欺骗自己意味着自己是知道真实信息的欺骗者,同时又是相信错误信息的被骗者。这样就造成一个似乎无法成立的悖论——自己在相信真实信息的同时又相信错误信息。解决这个问题的可能途径有几种。第一,个体一开始作为欺骗者,既意识到真实信念又意识到错误信念,然后慢慢说服自己相信错误信念,最终成为被骗者,只意识到错误信念(Kipp, 1980)。但这种方式不是严格意义上的“同时”。第二,把自我分成不同的部分,不同的意识层

面或者多个封闭的认知系统之间没有交流,它们分别持有相互矛盾的信念(Davidson, 1985; Kurzban & Aktipis, 2007)。但这种解释削弱了欺骗的成分,因为很难说明哪部分自我欺骗了哪部分自我。第三,个体在某些动机或情感的驱使下用有偏差的方式来处理信息,以致相信了错误的信念(Balceris, 2008; Mele, 1997; Thagard, 2007)。这种解释主张个体无需同时拥有真实和错误信念,也无需分解自我,只需在接触到真实信息的同时却用有偏差的解释方式以致产生错误信念,就算是自欺。但这种解释低估了欺骗者需要持有真实信念的重要性,因为一个人从真实信息中推出错误结论,只能算是单纯的推理错误,就像算错算术题一样;而且这种解释忽略了内隐反应,真实信息无论到达意识的哪个层面,都可能或多或少地影响个体行为(Bargh, 1997; Greenwald, 1997),显示出不同部分自我的存在。本文看待自欺定义的角度更倾向于把第三和第二种看法结合起来,认为自欺需要有不同部分的自我,在某些动机或情感的驱使下,某部分自我欺骗了另一部分的自我。

### 1.2 心理学中的自欺研究

**1.2.1 自我报告和生理、行为反应的不一致体现自欺** 心理学研究者通常创设一个操作定义来描述自欺。例如 Gur 和 Sackeim (1979)认为自欺应该符合如下 4 条原则: 1)个体必须持有两个互相矛盾的信念; 2)这两个信念是同时存在的; 3)个体意识不到其中一个信念; 4)个体的动机决定了哪个信念能被意识到(Gur & Sackeim, 1979)。他们要求被试辨别自己和他人的声音,并测量被试在听到声音时的皮肤电位反应。在前一个任务中收到负面反馈的被试倾向于把自己的声音判断为他人的,即不愿意承认自己的声音;但其皮肤电位对自己声音与他人声音的反应不同,即皮肤电可以正确辨别自己和他人的声音。而收到正面反馈的被试不会否认自己的声音。这表明受到负面反馈的个体可能拒绝接受一个失败的自己,因此在主观报告上,自欺地不承认自己的声音,但是在生理反应层面却可以正确辨认自己和他人的声音,符合研究者对自欺的定义。尽管此后有研究者对皮肤电位可以辨认自己的声音提出不同看法(Douglas & Gibbins, 1983),上述研究仍然成为了心理学上探讨自欺的经典之作。后来有研究者也通过主观报告和生理反应的不一致来考察自欺。他们用问卷测量个体对同性恋的态度,相比起可以接受同性恋的被试,那些在问卷中表现得十

分讨厌同性恋的被试，在观看关于同性恋的录像时，其生殖器的反应更大，表明他们实际上对同性恋很有感觉却自欺地认为自己讨厌同性恋(Adams, Wright, & Lohr, 1996)。

除了上述两份研究，人们用实验的方式来探讨自欺，似乎都集中在个体主观报告和实际行为的不一致上。例如，人们会做出一些被认为能够导致好结果的行为，尽管他们不承认自己有这样做。被试把双臂浸入冷水箱中，坚持到无法忍受为止。然后一半被试得知“医学发现忍耐时间越长个体的心脏越健康”，另一半被试得知“忍耐时间越短心脏越健康”。之后实验者再测试一次被试把手放入冷水箱中的忍耐时间。结果显示，相信忍耐时间与心脏健康有正相关的被试第二次的忍耐时间明显长于第一次，而相信忍耐时间与心脏健康有负相关的被试则恰恰相反。但所有被试都认为自己第一次与第二次的忍耐时间没有差异(Quattrone & Tversky, 1984)。类似地，被试完成一个考察反应速度的任务，需要把鼠标放在屏幕左边的起始位置上，当屏幕右边的红点出现时，以最快的速度把鼠标移动到红点上。完成此任务后，一半被试得知“移动速度越快智力越高”(快速组)，另一半被试则得到相反的指示语(慢速组)。之后被试再完成一次任务，快速组被试移动鼠标的速度比第一次完成任务时快，而慢速组两次的移动速度没有区别；同时所有被试都报告自己没有故意加快或减慢移动速度，都是以自己的最快速度来移动的(Sloman, Fernbach, & Haggmayer, 2010)。

在道德上的研究也表明，人们认为自己是善良的、利他的、公正的，但其实际行为并非如此，而个体自己还察觉不到。在不记名调查中，被试回答自己实际上会把5美元被试费中的多少捐给慈善机构，所有被试平均起来认为自己会捐出2.5美元，而推测别人只会捐1.8元；当实验室门口确实摆放有捐款箱时，箱里的款额显示平均每名被试只捐了1.5美元。由于是不记名调查，被试无需在回答问卷时故意吹嘘，因此很可能是被试自欺地高估了自己的慷慨程度(Epley & Dunning, 2000)。在另一项实验中，研究者让被试给自己和另一个人分配任务，其中一个任务有趣且有报酬，另一个枯燥且没报酬。在选用抛硬币的方法来分配任务的被试中，只有10%的人给对方分配了有报酬任务；但如果按照抛硬币的结果，自己或对方得到有报酬任务的概率应该是一半一半。然而，被试认为自己的分配是公

正的，在分配公证性的评分上高于那些不选用抛硬币方式的被试，这表明被试可能依靠抛硬币的方式自欺地认为自己进行了公平的分配(Batson, Kobryniewicz, Dinnerstein, Kampf, & Wilson, 1997)。

**1.2.2 有偏差的信息加工导致自欺** 还有一系列关于认知偏差的研究，也被用于描述自欺。Greenwald (1988)认为自欺可以发生在认知加工过程的任何一个阶段中。人们会把自己不想知道的信息像垃圾信件一样扔掉，会自动过滤掉或屏蔽具有威胁性的信息，会按照自己的意愿来解释接收到的信息，会根据实际情况有偏差地提取信息或重构记忆。首先，在搜索信息时，如果收集到的信息符合个体的预期和愿望，个体就会很快停止搜索；但如果先期收集到的信息不符合预期，个体就会花费更多时间去搜索，以便有更多机会得到符合自己愿望的信息(Ditto & Lopez, 1992)。如果搜索到的信息不符合自己的意愿，个体会自动质疑这些信息的有效性和真实性(Ditto, Munro, Apanovitch, Scepanisky, & Lockhart, 2003)。第二，个体在感知信息时，会出现自欺性的偏差。被试如果因为接触了一些难度很大的任务而感到很无助，他们会在雪花随机分布的雪景图中知觉出有意义的图案，借此增加控制感(Whitson & Galinsky, 2008)。个体喜欢把自己知觉为比实际上更有吸引力，100%的自己面孔与揉合了20%俊俏面孔元素的自己面孔相比，个体把后者知觉为与自己更相似(Epley & Whitchurch, 2008)。不管孩子实际上长得像谁，父母都倾向于把孩子知觉为长得像父亲，尤其是在亲子疑虑较高的情境中(Chang, Lu, Lee, Li, & Sui, 2010)。对于双眼竞争任务中同时呈现的正性词语和中性词语，充满自信的个体更多报告看到正性词语(Stark & Keating, 1991)。第三，个体在解释或评价信息时，会用自欺的方式诠释信息。吸烟者尽管知道吸烟严重影响肺部健康，但是他们通常认为自己不会得肺癌，肺癌是由基因决定的，自己的吸烟量不足以导致疾病，多吃维生素可以消除烟草的不良影响等(Dillard, McCaul, & Klein, 2006)。男性在解释异性的笑容和礼貌性动作时，会认为对方有挑逗意味，高估对方的性意图(Linton & Wiener, 2001)。第四，个体在回忆信息时，会自欺地重构记忆，比如回忆恋人的行为，分手前的回忆可能偏重于对方爱冒险、浪漫的一面，分手后的回忆可能就会变成对方是鲁莽、过于理想的(Gilbert, Pinel, Wilson, Blumberg, & Wheatley, 1998)。个体回忆往事时，也会因事件所

涉及的情绪和人物的不同而重构事件的发生时间(Lu & Chang, 2009)。个体对自己如何预测事件结果的记忆,会受到当前客观情况或真实结果的影响,他们会“事后聪明”般地觉得自己从一开始就对整件事情了如指掌(Pieters, Baumgartner, & Bagozzi, 2006)。

**1.2.3 过于正面的自我认识——自欺问卷** 除了小部分实验研究,大量关于自欺的研究都采用自欺问卷。这些问卷的逻辑是,自欺者往往从过于正面的角度看待自己,他们会极力否认威胁自我形象的信息。最早是 Sackeim 和 Gur (1979)的自欺问卷(Self-Deception Questionnaire, SDQ),包含 20 条题目,内容涉及威胁个体自我形象的描述,例如“你是否愚弄过自己”,“你是否讨厌过你父母”,“你是否对很有吸引力的异性产生过性冲动”等。个体在 7 点量表上评定自己曾经有过这样的想法的频率(1 代表极少,7 代表很经常)。自欺程度越高的个体,越防御和否认那些威胁自己的描述,总是评定自己很少有上述想法。每逢有题目被选了 1 或 2 就计 1 分,最后得分越高的个体越自欺(得分范围为 0~20 分)。人们的自欺得分与自尊呈正相关,与抑郁、神经症状程度呈负相关(Sackeim & Gur, 1979)。由于测量时是匿名的,被试的回答只有他自己才能知道,因此测量结果能够反映被试对自己真正的想法。与之相对,研究者还编制了欺骗他人问卷(Other-Deception Questionnaire, ODQ),测量个体为了在他人面前留下良好印象而否认违背社会规范的倾向。由于 SDQ 含有乱伦、性幻想等争议性较高的题目,后来 Paulhus (1991)把 SDQ 中一些与总体因素相关较低的题目删去,并增加一些非精神症状性的题目,编制了自欺性提升量表(Self-Deceptive Enhancement, SDE),题目包括“我总是知道我为什么喜欢某件事物”,“我从不后悔自己所做的决定”,“我能完全掌控自己的命运”等。他又在 ODQ 的基础上编制了印象管理量表(Impression Management, IM),与 SDE 共同组成社会赞许行为均衡量表(Balanced Inventory of Desirable Responding, BIDR)。后来他又编制了自欺性否定量表(Self-Deceptive Denial; Paulhus, 1999),题目包括“我从来都没有幸灾乐祸”,“我极少有性幻想”,“我所做的事情很少是单纯为了自己的”等,但由于此量表与 SDQ 一样存在争议性题目,以及它与印象管理量表相关较大,因此很少被应用。

由于存在自欺量表可供使用,长期以来研究者

把自欺当成一种反映个体差异的特质,通过量表得分来甄别自欺程度不同的个体。然而,自欺量表是否真正反映自欺,仍有待商榷。该量表的基本假设是个体肯定有过诸如“讨厌父母”这类不好的想法,或在潜意识中有这些想法,如果个体极力否认,那就是自欺。可是这个基本假设不一定正确,如果存在如圣人一般的人,他们从来没有过不好的想法,那么自欺量表就无法把他们与高度自欺的人区分开来,而且自欺量表也没有直接探测个体的潜意识,不能证实其假设。但由于量表使用方便,人们研究自欺时几乎都采用这个唯一的工具,因此以个体特质的方式来探讨自欺就成为了自欺研究的大方向。

### 1.3 自欺的作用

采用自欺问卷的研究通常探讨自欺对个体情绪、认知、决策、行为等的积极或消极作用。从好的方面说,自欺有助于提升主观幸福感(Erez et al., 1995; Taylor & Brown, 1988),压抑负性情绪(Robinson & Ryff, 1999),抑制反社会思维和行为(Otter & Egan, 2007)。自欺有利于人们提高自信和自我感觉的个人魅力(Linton & Wiener, 2001; Wrangham, 1999),使个体倾向于注意正面的信息(Stark & Keating, 1991),并且从积极的角度去看待自己及亲朋好友(Krebs & Denton, 1997; Murray & Holmes, 1997)。自欺可以帮助个体抵御负面信息,让个体在完成任务时不受威胁性信息的干扰(Paulhus & Suedfeld, 1988),或者在看到负面反馈时把注意力转向其它方面(Baumeister, 1996)。自欺还可以增强个体的控制感(Hagedorn, 1996; Morris, Sim, & Giroto, 1998),并使个体维持较高的道德标准,促使利他行为的出现(Lu & Chang, 2011a)。

从坏的方面说,自欺妨碍个体正确地认识自己及改正错误。由于看到考卷答案而获得高分的被试,自欺地认为分数高的原因是自己能力强,他们总是高估自己下一份考卷的成绩(Chance, Norton, Gino, & Ariely, 2011)。自欺者的焦虑敏感水平较低,他们较少选择任务取向的方式来解决,较多采用否认的策略,因此无助于问题解决(Werhun, Hons, & Cox, 1999)。自欺使个体减少努力从而影响其在学习和工作任务中的表现(Hirschfeld et al., 2008; Johnson, 1995; Lee & Klein, 2002; Martocchio & Judge, 1997)。例如,被试完成一个无法解决的拼图任务后,把字母组合成单词,自欺者受到挫败后就不愿意付出努力完成随后的任务,他们在组词任务上的表现比低自欺的人更差,而且他们把表现差归

咎于努力不够而非能力不足(Johnson, 1995)。高自欺者比低自欺者需要更长时间才能发现错误,而且会忽略那些错误,坚持用过时的策略。例如在博弈中,当输的次数越来越多时,低自欺者发现势头不对就会退出,而高自欺者还继续玩,结果输得更多(Peterson et al., 2003; Peterson, Driver-Linn, & DeYoung, 2002)。

#### 1.4 现有研究方向的不足

上述研究都是从“个体内”的角度探讨自欺,它们关注自欺的现象,造成自欺的因素,或自欺对人的影响。但这些研究都没有回答如下问题。既然自欺是认知过程出现了偏差,使人未能获得真实信息,那在人类进化过程中,自欺凭什么可以出现,并一直没被自然选择所淘汰呢?是因为自欺给个体带来的积极作用吗?如果一个心理机制在增加个体自信的同时妨碍了个体客观地认识自己,那么人类为什么没有发展出一个更完善的机制,使个体既有自信又能应付负面信息呢?要是只为了防御负面信息,以及呈现一个正面的自己,那么自欺根本不需要出现,只需要防御机制和乐观主义就够了。自欺使个体失去真实信息是有碍个体提高自身适应性的(van Leeuwen, 2007),它在人类进化过程中为什么会存在并被保留至今?

进化生物学家 Trivers 从人类在社群生活中的进化压力着手,从适应意义上思考自欺在进化过程中出现和存在的必要性,提出自欺产生的根本原因是个体需要更好地欺骗其他个体。这个观点尽管新颖,但从其提出到现在已经几十年,一直没有出现可行的理论假设,更不用说实证研究。本文根据上述观点的基本思想,为研究人类自欺提出一套理论假设和研究方法,在下文详细阐述。

## 2 从“人际间”的角度探讨自欺

自欺在人类进化过程中被自然选择保留下来,必定有其适应功能。在上述绝大部分的自欺研究中,自欺的主要功能是维持正面的自我形象,让个体保持愉快积极的心境。自欺阻挡或屏蔽了对自己有威胁性的信息,而这些信息往往反映个体真实能力或事件的真相。真实信息对个体生存很重要(Wallace, 1973),例如知道食物的确切位置可让个体免于挨饿,知道自己的真实水平可以促使个体不断改进和完善自己。自欺的个体把真实信息阻挡在意识之外,他们在维持正面情绪方面有得益,在获取真实信息方面有损失,其得益是否能高于损失,以致自欺没

有被自然选择所淘汰掉,这还是有个有待深入探究的问题。更关键的是,维持正面的自我形象不止自欺一种方式,个体还可以通过努力改善自己的方式来保持自尊和自信,阻挡威胁性的信息对目前来说可能有好处,但长远来说不利于个体正确认识自己(Robins & Beer, 2001),因此无论是保持正面情绪还是否认负面信息,都很难成为自欺被自然选择保留下来的主要原因。

在人类进化过程中,复杂的社群生活是人类大脑和认知能力迅猛发展的主要原因(Bailey & Geary, 2009; Dunbar, 1998)。合作和欺骗是群体生活的两大主题(Cosmides & Tooby, 1992)。欺骗是个体向对方传递错误信息,结果往往是对方利益受损而个体自己有所得益。欺骗如果被发现就会招致报复和惩罚。为了避免欺骗被识穿,欺骗者就发展出更完美的欺骗手段,比如在骗人之前,先让自己把将要传递的错误信息当作真实信息,也就是自欺。此时自欺的代价是失去真实信息,而其收益是成功骗过对方,获取骗人后的利益。如果这部分利益可以超过自欺的代价,自欺就会被自然选择保留下来,成为人类社会生活中具有适应意义的一种心理机制。

#### 2.1 欺骗与欺骗探测的“军备竞赛”

在自然界中,欺骗是求生和获得资源的一个普遍策略(Mitchell, 1986)。物种之间可以通过拟态来欺骗,蜂兰花长得像雌蜂吸引雄蜂来授粉,枯叶蝶长得像树叶逃避天敌,老鹰羽毛的斑纹酷似雀类的脸诱使小雀自投罗网(Gaskett, Winnick, & Herberstein, 2008; Negro, Bortolotti, & Sarasola, 2007)。同一物种的不同个体之间互相欺骗,用警报叫声吓走同伴,用迂回的路线引开竞争者,从而独占食物(Bugnyar & Kotrschal, 2004; Goodall, 1986; Held, Mendl, Devereux, & Byrne, 2002)。人类更是欺骗能手。婴儿无论是否饥饿,都会竭尽所能地哭,以便向父母展示自己的健壮并且获取父母更多的照顾(Lummaa, Vuorisalo, Barr, & Lehtonen, 1998);未能理解他人信念的3岁幼儿已经可以做出欺骗行为,例如通过增删脚印来导致别人寻找失败(Ruffman, Olson, Ash, & Keenan, 1993; Sodian, Taylor, Harris, & Perner, 1991);5岁的儿童在偷看了盒子里的玩具之后,可以假装自己没有偷看过,而且假装猜不到盒子里面是什么(Polak & Harris, 1999)。成人之间的欺骗每天都在发生,人们需要付出努力来控制自己才能做到不欺骗,如果脑力资源被剥夺,欺骗行为就更容易出现(Mead, Baumeister,

Gino, Schweitzer, & Ariely, 2009)。

相对的,探测欺骗是保障资源和避免利益受损的对策。欺骗对象不可能被动地等着受骗,而欺骗者由于承受着谎言可能被揭穿的压力,他们容易流露出一些暴露欺骗行为的线索,例如不自然的眼神、颤抖的声音、不自在的小动作、手心出汗等(Trivers, 2000),这些线索会被欺骗对象用来探测欺骗(Anderson, DePaulo, Ansfield, Tickle, & Green, 1999; DePaulo et al., 2003; Ekman & O'Sullivan, 1991)。人们可以通过神态和声音判断别人是否真心喜欢某件物品(DePaulo, Jordan, Irvine, Laser, 1982),靠分析面部表情(Frank & Ekman, 1997; Winston, Strange, O'Doherty, & Dolan, 2002)、小动作(Caso, Maricchiolo, Bonaiuto, Vrij, & Mann, 2006)、语速(DePaulo et al., 2003)、陈述内容的一致性(Vrij et al., 2008)来辨认撒谎者。人们还会注意并且记着欺骗者的面孔(Mealey, Daood, & Krage, 1996; Vanneste, Verplaetse, Hiel, & Braeckman, 2007)。

在社群生活中,人们对违反社会契约的欺骗行为也特别敏感。同样难度的推理题目,如果放在有潜在欺骗可能的情境中,人们的推理速度和准确度都会大大提高(Cosmides, 1989)。在一个有等级的群体系统中,通常会有一些公共契约或社会规范,如分配规则或行为禁忌等(Cummins, 1999)。如果社群成员不监控这些规则的实施,自私的个体就会违反这些规则来获得额外的利益,例如不付出劳动却蚕食群体利益,因此在长期进化过程中,人类对有可能造成自己利益损失的欺骗就特别敏感,而且一旦发现欺骗,就会狠狠地惩罚欺骗者。欺骗者为了避免欺骗被发现而招致攻击,就需要有更高明的欺骗策略。欺骗的技巧在不断进步,探测欺骗的手段也在进步,两者就像在进行“军备竞赛”,而自欺就是两者比拼的结果(Trivers, 2000),自欺是欺骗的终极形式,用于反欺骗探测(Lu & Chang, 2011b)。

自欺的个体把欺骗变为“诚实”,他们无需再担心欺骗被揭穿,因为他们根本意识不到自己在欺骗,各种能够泄露欺骗的言语和非言语线索都不复存在,对方想探测欺骗也无从着手(Trivers, 1985, 2000, 2009; von Hippel & Trivers, 2011)。Trivers认为个体在有意欺骗时,真实信息和错误信息都存在于意识中,个体一方面要努力保持镇定,不让自己表现出紧张,另一方面又要运用一些认知资源去压抑真实信息,这些行为都会让对方有机会探测出欺

骗;而个体自欺时,他们自己是意识不到真实信息的,他们的意识中只有错误信息,因此既没有紧张的困扰又不需要努力压制错误信息,此时的欺骗是个体把自己也当真的错误信息诚实地传递给别人。

## 2.2 研究自欺的情境

Trivers从1976年就开始提出自欺产生的根本原因,然而这个思想到目前为止还停留在理论阶段,没有发展出可验证的假设。本文沿着人际间欺骗的思路,提出欺骗探测压力是自欺产生的重要影响因素,并列存在不同欺骗探测压力的情境。有了这些情境,我们就可以比较个体在何时采用欺骗策略,在何时采用自欺策略,从而考察自欺是否为了更好地进行欺骗。由于自欺要付出损失真实信息的代价,在直接欺骗能够成功的情况下,个体没必要自欺;而当直接欺骗成功几率小时,自欺就发挥其作用(Lu & Chang, 2011b)。直接欺骗难于成功主要由于对方探测欺骗的能力高,在这种情况下,个体通过自欺来欺骗,就可以令对方无法探测到欺骗的存在。因此,自欺应该发生在对方欺骗探测能力比较高的时候,相反,如果对方的探测能力弱,个体就可以直接欺骗,而不需要担心欺骗被识穿。个体探测能力的强弱与其在群体生活中所处的位置有关。下文分别从地位高低、道德高低、人数众寡三个方面来阐述个体欺骗探测能力的强弱,进而推断欺骗者在面对哪些个体时会采用自欺策略。

根据社会占优理论(Social Dominance Theory, Cummins, 1999; 2005),一个等级系统中会形成一套潜在的交往规范,规定了处于不同地位的成员的行为界限,比如黑猩猩群体中,谁能坐在谁旁边,谁可以跟谁一起分享食物,谁能够为谁梳理毛发,谁能够首先获得配偶等等。高地位者拥有特权,是因为它们在群体形成之初已经通过竞争或打斗显示了自己的实力,确立了等级系统以及随之出现的分配规则,它们可以优先获得资源,包括食物和配偶。低地位者在武力上处于弱势,如果要获得额外资源,只能通过一些偷偷摸摸的方式,譬如隐瞒食物的存在,分散对方的注意力,或者遮掩自己的越级行为等(Bräuer, Call, & Tomasello, 2007; Bugnyar & Kotrschal, 2004; Hauser, 1998)。高地位者要探测这类欺骗从而保持其资源。探测违规行为的能力,高地位者好于低地位者。例如被试想像自己是监督员,对方是学生(高地位组),或者自己是学生对方是监督员(低地位组),对违背某个规定的作弊行为进行推理,也就是探测欺骗,高地位组的表现好于

低地位组(Cummins, 1999)。人类对地位低的欺骗者更为防范,例如更能记住地位低的欺骗者的面孔(Mealey et al., 1996)。因此,个体在欺骗地位比自己高的人时,成功的难度会增大,而且一旦欺骗被识破,就会招致高地位者的攻击和惩罚;此时如果采取自欺策略,就可以避免这些风险。

把地位等级系统借用到道德系统中,道德系统中地位高的人对违背道德规范的欺骗行为可能也有更好的探测能力。根据Kohlberg的道德发展理论,当道德水平发展到第四阶段时,个体会履行自己对社会的责任,用情感和良心来推动道德的执行,维持社会秩序,维护社会和群体的福利。高道德者擅于监控周围的人是否遵从道德,因为他们自己会表现出高尚的行为,这些行为有利于整个群体,例如捐钱给慈善机构,不随地扔垃圾等,如果低道德者没有这么做,就相当于侵占了高道德者所创造出来的好处,因此高道德者会时刻监控其他人。相反,低道德者没有做出太多高尚的行为来维护群体利益,没有被骗可言,也就不需要探测欺骗。人们评价他人行为时,道德是一个重要标准(Wojciszke, 1994),即便个体无意违反了公德,也会被认为是故意的(Leslie, Knobe, & Cohen, 2006)。因此人们在涉及道德的事情上都非常小心谨慎,如果做了任何不符合高道德标准的事情通常会试图隐瞒,尤其是在道德标准高的人面前。因此,人们在欺骗道德高的人时,容易感到比较大的探测压力,于是容易采用自欺的方式来进行欺骗。

当欺骗对象的人数众多时,欺骗者要承受的欺骗被探测的压力应该比只欺骗一个人时大,因为每个欺骗对象都是一份探测欺骗的力量。欺骗者为了表明自己向所有人说的话是真话,很有可能本来明知道自己在撒谎,后来也变得相信了自己的谎言(Tice, 1992)。当有观众存在时,个体的亲社会行为和合作行为会增加(Andreoni & Petrie, 2004; Burnham, 2003; Kurzban, DeScioli, & O'Brien, 2007; Piazza & Bering, 2008),欺骗行为会减少(Diener, Fraser, Beaman, & Kelem, 1976)。有时候不一定需要存在真正的观众,只需给个体营造一种被监察着的感觉,例如呈现类似眼睛的物体(Burnham & Hare, 2007)或者眼睛图案(Ernest-Jones, Nettle, & Bateson, 2011; Haley & Fessler, 2005; Oda, Niwa, Honma, & Hiraishi, 2011),个体的合作行为就会增加,违规行为会减少。由于存在平衡群体中不同成员利益的道德系统,以及经常监控别人是否遵守规

则的群体成员,个体需要压抑自私的想法,表现出受社会赞许的行为。如果自私的想法存在于意识中,却要时刻向别人显示自己的高道德水平,个体就要随时警惕欺骗是否会被揭穿,这样压力会很大。如果自私的想法被压抑到无意识中,也就是自欺,那么个体即使面对再多的观众,也能应付自如。

根据上述分析,要验证自欺是为了更好地欺骗他人这个观点,首先应该让欺骗者感觉到对方的探测能力比较强,也就是创设出欺骗探测压力高的情境,并且与探测压力低的情境相比较,考察个体在探测压力不同时,是否采取不同的欺骗策略。如果在探测压力较弱时,个体直接欺骗,而在探测压力较强时,个体表现出自欺,那就说明自欺的作用是应付欺骗探测,使得个体可以欺骗成功。

### 2.3 人际间和个体内自欺的关系

从人际间的角度看待自欺,其作用是让个体更有效地进行欺骗,而从个体内的角度看待自欺,其作用是帮助个体抵御威胁自我形象的信息,增强个体的自信心和控制感。无论从哪个角度进行解释,自欺都能为个体带来利益,弥补自欺的损失,那为什么说是人际间的欺骗而非个体自身的需要推动了自欺的产生呢?正如前文提到,如果仅仅是为了防御负面信息或提升自我,运用其它心理机制就足够,不需要额外发展出自欺这个新的机制。因此自欺最早是在人际间欺骗的压力下产生的,而它出现后,被借用到个体内的领域,帮助个体提升自我。就像嘴巴最初的适应功能是进食,后来也用于帮助发声和说话一样。

那么,进一步的问题是,个体为什么要防御负面信息和提升自我呢?因为维持正面的自我形象,增强自己的控制感等,最终都是为了在别人面前呈现一个能力强、道德高的自己,并增强自己的说服力,让别人按照自己的提议行事。假设在一个人们不评价别人能力和道德的社会中,能力强的人不获赞赏,道德高的人不受推崇,个体就没有必要维持正面的自我形象。再假设有这么一个社会,在那里个体的能力、道德根本不重要,只要是可以不断交配繁殖的人就是成功者,那里的人以做坏事为荣,尊重和崇拜可以想出自私的办法来利己的人,在那里利他行为是一种耻辱,人们都争着表现自己是个自私的人,在这样一个社会里,自欺性的自我提升是毫无意义的。因此,作用于自身的自欺,最终是为了让他人相信我是一个能力强、道德高的人,也就是用于欺骗他人的。

再进一步的问题是,为什么人类社会推崇能力好和道德高的个体呢?因为生物体是自私的(Dawkins, 1976; Alexander, 1987)。别人如果有能力找到食物并且愿意与我分享,对我来说是有利的,因此我希望我身边的人都又能干又替我着想。其他人也持有这种想法,因此整个社会就推崇能干和品德高的人。既然能力和道德只是表现给别人看的,为什么不直接欺骗?因为欺骗探测的存在使得个体需要更巧妙的欺骗方式。首先,能力很难假装,如果个体明知自己能力差还夸夸其谈,在要实战时就会露出马脚。但如果个体真心相信自己能力高,在实际做的时候就会干劲十足,勇于应付挑战,这样自我实现的预言就很有可能出现(Krebs, Denton, & Higgins, 1988; Snyder, 1984),让个体真的显示出能力高。其次,假装道德高容易让个体内心产生矛盾。当个体明知自己不愿与人合作却要表现出高尚和利他时,这种赤裸裸的表里不一就容易给个体造成情绪困扰,引起各种负面情绪(Emmons & King, 1988; O'connor et al., 2002)。另外个体从小所受的教育就是要争取成为能力好、道德高的人,以至于这种欺骗其他所有人的需要内化成个体对自己的要求,尽管个体不如自己所说的那么慷慨,他们还是真心地认为自己是慷慨大方的,而这种信念要通过自欺来达成(Essock-Vitale, McGuire, & Hooper, 1988; Epley & Dunning, 2000)。个体内自欺,其实就相当于欺骗对象众多的人际间欺骗,个体身边的人都是欺骗探测源。只要个体是生活在群体中,向外界呈现一个正面形象就是一项持续的任务,即欺骗所有人是时刻都在发生的,因此自欺策略就一直被运用着。

### 3 研究人际间自欺的方法

现有的自欺研究,无论是主观报告与生理、行为反应的不一致,还是问卷测量的自欺特质,或者是认知过程中的自欺性偏差,都没有把自欺放在人际间欺骗的情境中来考察。本文提出采用两次记忆的方式来研究自欺,这种方法可用于欺骗情境中,并且符合研究者从进化角度对自欺的定义。

#### 3.1 用两次记忆的方式来研究自欺

根据 Trivers (2000), von Hippel 和 Trivers (2011)对自欺的描述,个体在欺骗进行时意识不到真实信息,此时真实信息存在于无意识中,个体意识到的是错误信息,他把错误信息当作真实信息传递给对方,通过自欺来完成欺骗。本文提出用记忆来直观

地考察这类自欺。由于记忆可分为外显记忆和内隐记忆,两种记忆方式往往可以提取到不同的信息,个体自欺时,内隐记忆的内容很可能是真实信息而外显记忆却显示错误信息。记忆内容的可塑造范围很大,如果在欺骗完成后,个体在不需再自欺的情况下,可用外显记忆的方式提取真实信息,那么他就既可以获得欺骗的好处,又不用承担失去真实信息的损失。

假设低地位者被高地位者问及食物在哪里,他想隐瞒食物的位置,欺骗高地位者说不知道,如果有意识地隐瞒,可能会被识穿而招致攻击。如果他当时真的忘记食物在哪里,就可以理直气壮地说不知道,不需要担心欺骗被识破,因为个体已经自欺了,连他自己也不知道自己在进行欺骗。事后他要是能够想起来食物在哪里,他就可以独享食物。在需要隐瞒时想不起真实信息,在不需要隐瞒时能记起来,这种自欺情境可以通过两次记忆测试来模拟。记忆测试的编码阶段就相当于个体获取有用信息的阶段,这些信息对个体来说是重要的,代表着个体不愿意与其他人分享的资源、或是获取这些资源的渠道、又或是个体要深藏而不能向别人表露的想法和行为。

第一次记忆测试应该放在个体持有欺骗动机的情境中进行。例如个体拥有关于某些重要资源的信息,或者他做了一些违背等级规定的行为,这些都需要向别人隐瞒,尤其是地位高的对象。但如果是故意隐瞒,一旦被对方发现,就会遭到惩罚;此时一种有效的隐瞒方式就是个体连自己也提取不到那些需要被隐瞒的信息,这样他就能既诚实地面对别人,又不把真实信息传递出去。由于个体在意识上毫无保留地把自己能够想起来的信息都告诉别人,表现得“诚实可信”,他们觉察不到自己在隐瞒或欺骗,也就是做到了自欺。而实际上,那部分真实信息仍然被保留在无意识中,只是没有被个体意识到。要模拟这种自欺方式,可以在实验中诱发个体的欺骗动机,使其不愿意与他人分享在编码阶段获得的信息,或者想要向他人隐瞒那些被编码的信息。然后测量个体在与别人分享信息时的记忆表现。如果个体表现出自欺,也就是在欺骗动机存在时他忘记了某些真实信息,其记忆成绩就会差于在欺骗动机不存在时。当然,欺骗动机的存在占用了个体一些认知资源去犹豫是否要进行欺骗,本来就有可能影响到个体在其它认知任务(比如记忆)上的表现,因此需要比较个体在欺骗不同对象时的记忆



表现。如果记忆变差只发生在个体欺骗某些人时，而不是在所有欺骗情境中，那就说明记忆变差主要不是由于认知资源被占用。根据我们的假设，自欺应该发生在对方的欺骗探测能力比较高的时候，如果对方的欺骗探测能力很差，个体不担心欺骗被识破，那么直接欺骗就可以了。当欺骗对象的探测能力比较高时，例如对方地位高或人数众多，个体不敢直接欺骗，于是他们会表现得诚实，但这种“诚实”可以是个体意识不到自己正在进行欺骗，也就是通过记忆变差来隐瞒信息。当欺骗对象的探测能力弱时，个体的记忆不需要变差，他们会进行直接欺骗，在记得真实信息的情况下却不告诉对方。因此，如果自欺与否受对方欺骗探测能力的影响，在欺骗动机存在时，个体在高探测力对象面前的回忆表现应该差于个体面对低探测力的对象时。

第二次记忆测试应该放在欺骗动机消失后的情境中进行。当个体不需要再面对欺骗对象时，那些有用的真实信息如果能够浮现到意识中，就可以为个体带来好处。由于个体可能会回忆起一些之前在欺骗动机存在时未被提取到的信息，此时测量个体的记忆表现，其记忆成绩应该好于当初在欺骗动机存在时。对于个体不想与别人分享的真实信息，用自欺的方式向别人隐瞒，个体自己以失去真实信息为代价，但暂时是没有欺骗成功的获益的，只有当个体可以重新获得真实信息时，其欺骗行为才能为自己带来好处。而重新获得的真实信息，就是第二次记忆测试中期望测量到的那部分记忆成绩的提高。对于个体想要藏起来的违规行为或事实真相，在欺骗成功时个体就已经避免了受惩罚，获得了欺骗的好处，如果个体意识不到这些行为或想法对个体生存来说没有代价，那么这些信息就有可能永远不被意识到，如果存在代价，个体在不需要欺骗时可以意识到这些信息，就能将代价减少。

采用两次记忆的方式来研究自欺，一来因为这种程序可以比较直观地表现个体自欺时真实信息存在于无意识中，未被提取出来；而通过比较两次回忆的差异，可以反映出第一次回忆时储存在无意识中的信息，这样既绕过了测量无意识的困难，又揭示了自欺的存在。二来因为记忆作为人类社群生活中一种不可或缺的认知能力，经过进化过程的长期选择，它具备很多有适应性的特点。首先，记忆系统能够帮助个体记住与生存相关的知识，提高个体的生存几率。在进化过程中，人类祖先需要记住哪里有食物哪里有猛兽出没，记住哪些食物有毒哪

些食物美味等，这对个体的生存非常重要。人们可以更好地记住合作者和欺骗者的长相(Mehl & Buchner, 2008)，记住在与生存相关的情境中被编码的信息(Nairne, Pandeirada & Thompson, 2008; Nairne, Thompson, & Pandeirada, 2007)，这反映了被自然选择保留下来的记忆特点，是有助于提高个体的生存适应性的。第二，事后认识偏差(Hindsight Bias)显示人们能够按照现实结果来重构自己的记忆(Fischhoff, 1975)，指导以后的行为(Pieters et al., 2006)。记忆具有疗伤功能，不愉快经历所带来的负面情绪强度会随着时间推移而下降，但愉快经历的情绪强度在记忆中稳定不变(Ritchie et al., 2006; Skowronski, Gibbons, Vogl, & Walker, 2004; Wood & Conway, 2006)，记忆也可以通过改变人们的心理感受来减低牵涉亲人的不愉快事件对当前生活的影响(Lu & Chang, 2009)。这些记忆偏差都有利于让个体维持积极的情绪状态，更好地应对现在和未来的生活。第三，定向遗忘范式表明人们可以抑制提取不想记起的内容(Barnier et al., 2007; Bjork & Bjork, 1996)，尤其是负性信息(Myers, Brewin, & Power, 1998)，而且遗忘是因为提取遭到压抑，而非对信息编码的不充分(Bjork & Bjork, 1996)，因为个体对需要遗忘和不需要遗忘的内容，在内隐记忆上没有区别(David & Brown, 2003)。这表明个体能够回忆起哪些内容，遗忘哪些内容，是受其记忆动机影响的。第四，外显和内隐记忆的研究表明个体通过回忆提取出来的信息只是其所储存的内容的一部分，个体有可能把已知的信息压抑到无意识中。上述四个记忆的特点为个体自欺提供了有利条件。

人类记忆系统对信息进行编码、储存、提取和重构时所表现出来的这些特点，是自然选择机制筛选出来的，有利于提高个体适应性。对个体无用的信息被压抑，对个体有用的信息被提取。如果是同一项信息，无用和有用只是相对于情境来说的，譬如在自欺的例子中，在欺骗进行时，真实信息对个体来说是无用的，在欺骗过后，真实信息对个体来说是有用的，那么记忆系统也许能够根据不同情境来改变其对真实信息的处理，帮助个体实现自欺。

### 3.2 自欺策略的启动

如果把意识分为不同层次，主观报告的信息表明个体意识到，而生理反应和行为受无意识影响(Bargh, 1997)，那么自欺有可能是持有真实信息的无意识欺骗了持有错误信息的有意识(Thagard,

2007)。当个体连自己都不知道自己正在欺骗时,他是如何感受到欺骗探测压力的呢?本文假设,在无意识中有一个机制,负责启动自欺策略。这个机制可以感受到不同欺骗对象所带来的探测压力,并且迅速计算直接欺骗成功的几率以及自欺的收益和代价,最后决定是否启动自欺策略。这个机制对关于欺骗对象的信息应该是很敏感的。欺骗对象所透露出来的各种信号,如眼神、表情、声音、话语、打扮等,都应该会引起这个机制对欺骗对象探测能力的判断。而这种判断在长期自然选择的压力下,可能已被塑造成无意、自动、快速的,就像条件反射一样。不少研究表明个体可以对别人进行快速分类,从而做出相应的社交行为(e.g., Devine, 1989; Jost & Banaji, 1994)。而且认知上的研究已表明,由无意识做出的决定通常是最好的,尤其是当决策需要多项信息时,暂时不主动思考或睡一觉,把计算留给无意识来完成,个体最后往往可以做出最佳选择(Dijksterhuis, Bos, Nordgren, & van Baaren, 2006)。因此本文推测,无意识可以迅速觉察对方的欺骗探测能力,并暗中计算得失,从而决定是否启动自欺策略。另外,个体所感觉到的压力,有可能与对方的实际探测能力毫无关系,只要对方被个体分到探测能力高的类别中,那个无意识机制就自动启用自欺策略。当然这些推测还需要更多实验来严谨地证明。

#### 4 研究自欺的意义

自欺有时带来的欺骗可以帮助人们提升自信心(Paulhus & John, 1998),完成本来以为不可能的事情(Krebs & Denton, 1997),促进个体与他人的合作(Lu & Chang, 2011a),这些都是其好的方面。然而,自欺有时带来的欺骗使个体陷入危害社会的状态而不自知,例如公司决策者低估自己的项目对环境所造成的破坏,有时个体的自欺信念导致整个集体的自欺以致产生严重的社会影响,例如希特勒的纳粹主义(Triandis, 2009)。因此研究自欺产生的原因有助于利用自欺的有利方面以及防止自欺的不良后果。由于自欺有时像反射一样,个体未必意识到行为背后的原因(Shapiro, 1996),因此很难对自欺者的意图性做道德判断,因此更应该注意容易让自欺出现的条件,从而预防会带来重大不良后果的自欺。从进化的视角看待自欺,解释了自欺产生的根本原因,也就是反欺骗探测,这样有助于研究者从这个原因出发,推导出有可能使自欺产生的条件

和情境,让人们可以更好地理解和利用自欺。

#### 参 考 文 献

- Adams, H. E., Wright, L. W., Jr., & Lohr, B. A. (1996). Is homophobia associated with homosexual arousal? *Journal of Abnormal Psychology, 105*, 440-445.
- Alexander, R. D. (1987). *The biology of moral systems*. New York: Aldine.
- Anderson, D. E., DePaulo, B. M., Ansfield, M. E., Tickle, J. J., & Green, E. (1999). Beliefs about cues to deception: Mindless stereotypes or untapped wisdom. *Journal of Nonverbal Behavior, 23*, 67-89.
- Andreoni, J., & Petrie, R. (2004). Public goods experiments without confidentiality: A glimpse into fund-raising. *Journal of Public Economics, 88*, 1605-1623.
- Bailey, D. H., & Geary, D. C. (2009). Hominid brain evolution: Testing climatic, ecological, and social competition models. *Human Nature, 20*, 67-79.
- Balcetis, E. (2008). Where the motivation resides and self-deception hides: How motivated cognition accomplishes self-deception. *Social and Personality Psychology Compass, 2*, 361-381.
- Bargh, J. A. (1997). The automaticity of everyday life. In R. S. Wyer, Jr. (Ed.), *The automaticity of everyday life* (pp. 1-61). Mahwah, NJ: Lawrence Erlbaum.
- Barnier, A. J., Conway, M. A., Mayoh, L., Speyer, J., Avizmil, O., & Harris, C. B. (2007). Directed forgetting of recently recalled autobiographical memories. *Journal of Experimental Psychology: General, 136*, 301-322.
- Batson, C. D., Kobryniewicz, D., Dinnerstein, J. L., Kampf, H. C., & Wilson, A. D. (1997). In a very different voice: Unmasking moral hypocrisy. *Journal of Personality and Social Psychology, 72*, 1335-1348.
- Baumeister, R. F. (1996). Self-regulation and ego threat: Motivated cognition, self deception, and destructive goal setting. In P. M. Gollwitzer & J. A. Bargh (Eds.), *The psychology of action: Linking cognition to behavior* (pp. 27-47). New York: The Guilford Press.
- Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness and Cognition, 5*, 176-196.
- Bräuer, J., Call, J., & Tomasello, M. (2007). Chimpanzees really know what others can see in a competitive situation. *Animal Cognition, 10*, 439-448.
- Bugnyar, T., & Kotrschal, K. (2004). Leading a conspecific away from food in ravens (*Corvus corax*)? *Animal Cognition, 7*, 69-76.
- Burnham, T. C. (2003). Engineering altruism: A theoretical and experimental investigation of anonymity and gift giving. *Journal of Economic Behavior & Organization, 50*, 133-144.
- Burnham, T. C., & Hare, B. (2007). Engineering cooperation: Does involuntary neural activation increase public goods contributions. *Human Nature, 18*, 88-108.
- Caso, L., Maricchiolo, F., Bonaiuto, M., Vrij, A., & Mann, S. (2006). The impact of deception and suspicion on different hand movements. *Journal of Nonverbal Behavior, 30*, 1-19.
- Chance, Z., Norton, M. I., Gino, F., & Ariely, D. (2011). Temporal view of the costs and benefits of self-deception. *Proceedings of National Academic Society, 108*, 15655-15659.
- Chang, L., Lu, H. J., Lee, L. N. Y., Li, T., & Sui, S. S. (2010). Patrilocal residence and father-child resemblance beliefs in

- relation to paternal investment. *Parenting: Science and Practice*, 10, 274–285.
- Cosmides, L. (1989). The logic of social exchange: Has natural selection shaped how humans reason? Studies with the Wason selection task. *Cognition*, 31, 187–276.
- Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In H. J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 163–228). New York: Oxford University Press.
- Cummins, D. (2005). Dominance, status, and social hierarchies. In D. M. Buss (Ed.), *The Handbook of evolutionary psychology* (pp. 676–697). Hoboken, NJ: John Wiley & Sons.
- Cummins, D. (1999). Cheater detection is modified by social rank: The impact of dominance on the evolution of cognitive functions. *Evolution and Human Behavior*, 20, 229–248.
- David, D., & Brown, R. J. (2003). The impact of different directed forgetting instructions on implicit and explicit memory: New evidence from a modified process dissociation procedure. *Quarterly Journal of Experimental Psychology: Section A*, 56, 211–231.
- Davidson, D. (1985). Deception and division. In E. LePore & B. P. McLaughlin (Eds.), *Actions and events: Perspectives on the philosophy of Donald Davidson* (pp. 138–148). Oxford, England: Blackwell.
- Dawkins, R. (1976). *The selfish gene*. Oxford: Oxford University Press.
- DePaulo, B. M., Jordan, A., Irvine, A., & Laser, P. S. (1982). Age changes in the detection of deception. *Child Development*, 53, 701–709.
- DePaulo, B. M., Lindsay, J. J., Malone, B. E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin*, 129, 74–118.
- Devine, P. G. (1989). Stereotypes and prejudice: Their automatic and controlled components. *Journal of Personality and Social Psychology*, 56, 5–18.
- Diener, E., Fraser, S. C., Beaman, A. L., & Kelem, R. T. (1976). Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology*, 33, 178–183.
- Dijksterhuis, A., Bos, M. W., Nordgren, L. F., & van Baaren, R. B. (2006). On making the right choice: The deliberation-without-attention effect. *Science*, 311, 1005–1007.
- Dillard, A. J., McCaul, K. D., & Klein, W. M. P. (2006). Unrealistic optimism in smokers: Implications for smoking myth endorsement and self-protective motivation. *Journal of Health Communication*, 11(s1), 93–102.
- Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63, 568–584.
- Ditto, P. H., Munro, G. D., Apanovitch, A. M., Scepansky, J. A. & Lockhart, L. K. (2003). Spontaneous skepticism: The interplay of motivation and expectation in responses to favorable and unfavorable medical diagnoses. *Personality and Social Psychology Bulletin*, 29, 1120–1132.
- Douglas, W., & Gibbins, K. (1983). Inadequacy of voice recognition as a demonstration of self-deception. *Journal of Personality and Social Psychology*, 44, 589–592.
- Dunbar, R. I. M. (1998). The social brain hypothesis. *Evolutionary Anthropology*, 6, 178–190.
- Ekman, P., & O'Sullivan, M. (1991). Who can catch a liar? *American Psychologist*, 46, 913–920.
- Emmons, R. A., & King, L. A. (1988). Conflict among personal strivings: Immediate and long-term implications for psychological and physical well-being. *Journal of Personality and Social Psychology*, 54, 1040–1048.
- 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, 861–875.
- Epley, N., & Whitchurch, E. (2008). Mirror, mirror on the wall: Enhancement in self recognition. *Personality and Social Psychology Bulletin*, 34, 1159–1170.
- Erez, A., Johnson, D. E., & Judge, T. A. (1995). Self-deception as a mediator of the relationship between dispositions and subjective well-being. *Personality and Individual Difference*, 19, 597–612.
- Ernest-Jones, M., Nettle, D., & Bateson, M. (2011). Effects of eye images on everyday cooperative behavior: A field experiment. *Evolution and Human Behavior*, 32, 172–178.
- Essock-Vitale, S. M., McGuire, M. T., & Hooper, B. (1988). Self-deception in social-support networks. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 200–211). Englewood Cliffs, NJ: Prentice-Hall.
- Festinger, L., & Carlsmith, J. M. (1959). Cognitive consequences of forced compliance. *Journal of Abnormal and Social Psychology*, 58, 203–210.
- Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 288–299.
- Frank, M. G., & Ekman, P. (1997). The ability to detect deceit generalizes across different types of high-stake lies. *Journal of Personality and Social Psychology*, 72, 1429–1439.
- Gaskett, A. C., Winnick, C. G., & Herberstein, M. E. (2008). Orchid sexual deceit provokes ejaculation. *American Naturalist*, 171, 206–212.
- Gilbert, D. T., Pinel, E. C., Wilson, T. D., Blumberg, S. J., & Wheatley, T. P. (1998). Immune neglect: A source of durability bias in affective forecasting. *Journal of Personality and Social Psychology*, 75, 617–638.
- Goodall, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, Mass: Belknap Press of Harvard University Press.
- Greenwald, A. G. (1988). Self-knowledge and self-deception. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 113–131). Englewood Cliffs, NJ: Prentice-Hall.
- Greenwald, A. G. (1997). Self-knowledge and self-deception: Further consideration. In M. S. Myslobodsky (Ed.), *The mythomanias: The nature of deception and self-deception* (pp. 51–72). Mahwah, NJ: Lawrence Erlbaum Associates.
- Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology*, 74, 1464–1480.
- Gur, R. C., & Sackeim, H. A. (1979). Self-deception: A concept in search of a phenomenon. *Journal of Personality and Social Psychology*, 37, 147–169.
- Hagedorn, J. W. (1996). Happiness and self-deception: An old question examined by a new measure of subjective well-being. *Social Indicators Research*, 38, 139–160.
- Haley, K. J., & Fessler, D. M. T. (2005). Nobody's watching?

- Subtle cues affect generosity in an anonymous economic game. *Evolution and Human Behavior*, 26, 245–256.
- Hauser, M. D. (1998). Games primates play. *Discover*, 19, 48–53.
- Held, S., Mendl, M., Devereux, C., & Byrne, R. W. (2002). Foraging pigs alter their behaviour in response to exploitation. *Animal Behaviour*, 64, 157–166.
- Hirschfeld, R. R., Thomas, C. H., & McNatt, D. B. (2008). Implications of self-deception for self-reported intrinsic and extrinsic motivational dispositions and actual learning performance: A higher order structural model. *Educational and Psychological Measurement*, 68, 154–173.
- Johnson, E. A. (1995). Self-deceptive coping: Adaptive only in ambiguous contexts. *Journal of Personality*, 63, 759–791.
- Jost, J. T., & Banaji, M. R. (1994). The role of stereotyping in system-justification and the production of false consciousness. *British Journal of Social Psychology*, 33, 1–27.
- Kipp, D. (1980). On self-deception. *Philosophical Quarterly*, 30, 305–317.
- Krebs, D. L., & Denton, K. (1997). Social illusions and self-deception: The evolution of biases in person perception. In J. A. Simpson & D. T. Kenrick (Eds.), *Evolutionary social psychology* (pp. 21–48). Mahwah, NJ: Erlbaum.
- Krebs, D., Denton, K., & Higgins, N. (1988). On the evolution of self-knowledge and self-deception. In K. MacDonald (Ed.), *Sociobiological perspectives on human development* (pp. 103–139). New York: Springer-Verlag.
- Kurzban, R., & Aktipis, C. A. (2007). Modularity and the social mind. *Personality and Social Psychology Review*, 11, 131–149.
- Kurzban, R., DeScioli, P., & O'Brien, E. (2007). Audience effects on moralistic punishment. *Evolution and Human Behavior*, 28, 75–84.
- Lee, S., & Klein, H. J. (2002). Relationships between conscientiousness, self-efficacy, self-deception, and learning over time. *Journal of Applied Psychology*, 87, 1175–1182.
- Leslie, A. M., Knobe, J. & Cohen, A. (2006). Acting intentionally and the side-effect effect: Theory of mind and moral judgment. *Psychological Science*, 17, 421–427.
- Linton, D. K., & Wiener, N. I. (2001). Personality and potential conceptions: Mating success in a modern Western male sample. *Personality and Individual Differences*, 31, 675–688.
- Lu, H. J., & Chang, L. (2009). Kinship effect on subjective temporal distance of autobiographical memory. *Personality and Individual Differences*, 47, 595–598.
- Lu, H. J., & Chang, L. (2011a). The association between self-deception and moral self-concept as functions of self-consciousness. *Personality and Individual Differences*, 51, 845–849.
- Lu, H. J., & Chang, L. (2011b). Self-deceive to countermeasures detection. *Behavioral and Brain Sciences*, 34, 33.
- Lummaa, V., Vuorisalo, T., Barr, R. G., & Lehtonen, L. (1998). Why cry? Adaptive significance of intensive crying in human infants. *Evolution and Human Behavior*, 19, 193–202.
- Martocchio, J. J., & Judge, T. A. (1997). Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy. *Journal of Applied Psychology*, 82, 764–773.
- Mead, N. L., Baumeister, R. F., Gino, F., Schweitzer, M. E., & Ariely, D. (2009). Too tired to tell the truth: Self-control resource depletion and dishonesty. *Journal of Experimental Social Psychology*, 45, 594–597.
- Mealey, L., Daoood, C., & Krage, M. (1996). Enhanced memory for faces of cheaters. *Ethology and Sociobiology*, 17, 119–128.
- Mehl, B., & Buchner, A. (2008). No enhanced memory for faces of cheaters. *Evolution and Human Behavior*, 29, 35–41.
- Mele, A. R. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20, 91–102.
- Mitchell, R. W. (1986). A framework for discussing deception. In R. W. Mitchell & N. S. Thompson (Eds.), *Deception, perspectives on human and nonhuman deceit* (pp. 3–40). Albany: State University of New York Press.
- Morris, M. W., Sim, D. L. H., & Giroto, V. (1998). Distinguishing sources of cooperation in the one-round Prisoner's Dilemma: Evidence for cooperative decisions based on the illusion of control. *Journal of Experimental Social Psychology*, 34, 494–512.
- Munn, C. A. (1986). Birds that “cry wolf”. *Nature*, 319, 143–145.
- Murray, S. L., & Holmes, J. G. (1997). A leap of faith? Positive illusions in romantic relationships. *Personality and Social Psychology Bulletin*, 23, 586–604.
- Myers, L. B., Brewin, C. R., & Power, M. J. (1998). Repressive coping and the directed forgetting of emotional material. *Journal of Abnormal Psychology*, 107, 141–148.
- Nairne, J. S., Pandeirada, J. N. S., & Thompson, S. R. (2008). Adaptive memory: The comparative value of survival processing. *Psychological Science*, 19, 176–180.
- Nairne, J. S., Thompson, S. R., & Pandeirada, J. N. S. (2007). Adaptive memory: Survival processing enhances retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 263–273.
- Negro, J. J., Bortolotti, G. R., & Sarasola, J. H. (2007). Deceptive plumage signals in birds: Manipulation of predators or prey? *Biological Journal of the Linnean Society*, 90, 467–477.
- Nesse, R., & Lloyd, A. (1992). The evolution of psychodynamic mechanisms. In J. H. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind* (601–624). New York: Oxford University Press.
- O'Connor, K. M., De Dreu, C. K. W., Schroth, H., Barry, B., Lituchy, T. R., & Bazerman, M. H. (2002). What we want to do versus what we think we should do: An empirical investigation of intrapersonal conflict. *Journal of Behavioral Decision Making*, 15, 403–418.
- Oda, R., Niwa, Y., Honma, A., & Hiraishi, K. (2011). An eye-like painting enhances the expectation of a good reputation. *Evolution and Human Behavior*, 32, 166–171.
- Otter, Z., & Egan, V. (2007). The evolutionary role of self-deceptive enhancement as a protective factor against antisocial cognitions. *Personality and Individual Differences*, 43, 2258–2269.
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver, L. S. Wrightsman & F. M. Andrews (Eds.), *Measures of personality and social psychological attitudes* (pp. 17–59). New York: Academic Press.
- Paulhus, D. L. (1999). BIDR Self-Deceptive Denial (SDD). Retrieved from <http://neuron4.psych.ubc.ca/~dpaulhus/research/SDR/downloads/MEASURES/SDD.htm>
- Paulhus, D. L., & John, O. P. (1998). Egoistic and moralistic biases in self-perception: The interplay of self-deceptive

- styles with basic traits and motives. *Journal of Personality*, 66, 1025–1060.
- Paulhus, D. L., & Suedfeld, P. (1988). A dynamic complexity model of self-deception. In J. S. Lockard & D. L. Paulhus (Eds.), *Self-deception: An adaptive mechanism?* (pp. 132–145). Englewood Cliffs, NJ: Prentice-Hall.
- Peterson, J. B., DeYoung, C. G., Driver-Linn, E., Séguin, J. R., Higgins, D. M., Arseneault, L., et al. (2003). Self-deception and failure to modulate responses despite accruing evidence of error. *Journal of Research in Personality*, 37, 205–223.
- Peterson, J. B., Driver-Linn, E., & DeYoung, C. G. (2002). Self-deception and impaired categorization of anomaly. *Personality and Individual Differences*, 33, 327–340.
- Piazza, J., & Bering, J. M. (2008). Concerns about reputation via gossip promote generous allocations in an economic game. *Evolution and Human Behavior*, 29, 172–178.
- Pieters, R., Baumgartner, H., & Bagozzi, R. (2006). Biased memory for prior decision making: Evidence from a longitudinal field study. *Organizational Behavior and Human Decision Processes*, 99, 34–48.
- Polak, A., & Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology*, 35, 561–568.
- Quattrone, G. A., & Tversky, A. (1984). Causal versus diagnostic contingencies: On self-deception and on the voter's illusion. *Journal of Personality and Social Psychology*, 46, 237–248.
- Ritchie, T. D., Skowronski, J. J., Wood, S. E., Walker, W. R., Vogl, R. J., & Gibbons, J. A. (2006). Event self-importance, event rehearsal, and the fading affect bias in autobiographical memory. *Self and Identity*, 5, 172–195.
- Robins, R. W., & Beer, J. S. (2001). Positive illusions about the self: Short-term benefits and long-term costs. *Journal of Personality and Social Psychology*, 80, 340–352.
- Robinson, M. D., & Ryff, C. D. (1999). The role of self-deception in perceptions of past, present, and future happiness. *Personality and Social Psychology Bulletin*, 25, 596–608.
- Ruffman, T., Olson, D. R., Ash, T., & Keenan, T. (1993). The ABCs of deception: Do young children understand deception in the same way as adults. *Developmental Psychology*, 29, 74–87.
- Sackeim, H. A., & Gur, R. C. (1979). Self-deception, other-deception, and self-reported psychopathology. *Journal of Consulting and Clinical Psychology*, 47, 213–215.
- Shapiro, D. (1996). On the psychology of self-deception. *Social Research*, 63, 785–801.
- Skowronski, J. J., Gibbons, J. A., Vogl, R. J., & Walker, W. R. (2004). The Effect of social disclosure on the intensity of affect provoked by autobiographical memories. *Self and Identity*, 3, 285–309.
- Sloman, S. A., Fernbach, P. M., & Haggmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, 115, 268–281.
- Snyder, M. (1984). When belief creates reality. *Advances in Experimental Social Psychology*, 18, 247–305.
- Sodian, B., Taylor, C., Harris, P. L., & Perner, J. (1991). Early deception and the child's theory of mind: False trails and genuine markers. *Child Development*, 62, 468–483.
- Stark, J. E., & Keating, C. K. (1991). Self-deception and its relationship to success in competition. *Basic and Applied Social Psychology*, 12, 145–155.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103, 193–210.
- Thagard, P. (2007). The moral psychology of conflicts of interest: Insights from affective neuroscience. *Journal of Applied Philosophy*, 24, 367–380.
- Tice, D. M. (1992). Self-concept change and self-presentation: The looking glass self is also a magnifying glass. *Journal of Personality and Social Psychology*, 63, 435–451.
- Triandis, H. C. (2009). *Fooling ourselves: Self-deception in politics, religion, and terrorism*. Westport, Conn.: Praeger Publishers.
- Trivers, R. (1976). Foreword. In R. Dawkins (Ed.), *The Selfish Gene*. NY: Oxford University Press.
- Trivers, R. (1985). Deceit and self-deception. In R. Trivers (Ed.), *Social Evolution* (pp. 395–420). Menlo Park: Benjamin/Cummings.
- Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of New York Academy of Sciences*, 907, 114–131.
- Trivers, R. (2009). Deceit and self-deception. In P. Kappeler & J. Silk (Eds.), *Mind the Gap* (pp. 373–393). New York: Springer-Verlag.
- Vanneste, S., Verplaetse, J., van Hiel, A., & Braeckman, J. (2007). Attention bias toward noncooperative people. A dot probe classification study in cheating detection. *Evolution and Human Behavior*, 28, 272–276.
- van Leeuwen, D. S. N. (2007). The spandrels of self-deception: Prospects for a biological theory of a mental phenomenon. *Philosophical Psychology*, 20, 329–348.
- von Hippel, W., & Trivers, R. (2011). The evolution and psychology of self-deception. *Behavioral and Brain Sciences*, 34, 1–16.
- Vrij, A., Mann, S. A., Fisher, R. P., Leal, S., Milne, R., & Bull, R. (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. *Law and Human Behavior*, 32, 253–265.
- Wallace, B. (1973). Misinformation, fitness, and selection. *The American Naturalist*, 107, 1–7.
- Werhun, C. D., Hons, B. A., & Cox, B. J. (1999). Levels of anxiety sensitivity in relation to repressive and self-deceptive coping styles. *Journal of Anxiety Disorders*, 13, 601–609.
- Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science*, 322, 115–117.
- Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2002). Automatic and intentional brain responses during evaluation of trustworthiness of faces. *Nature Neuroscience*, 5, 277–283.
- Wojciszke, B. (1994). Multiple meanings of behavior: Construing actions in terms of competence or morality. *Journal of Personality and Social Psychology*, 67, 222–232.
- Wood, W. -J., & Conway, M. (2006). Subjective impact, meaning making, and current and recalled emotions for self-defining memories. *Journal of Personality*, 74, 811–846.
- Wrangham, R. (1999). Is military incompetence adaptive? *Evolution and Human Behavior*, 20, 3–17.

## Self-Deception: Deceiving Yourself to Better Deceive Others

LU Hui-Jing

*(Department of Applied Social Sciences, The Hong Kong Polytechnic University, Hong Kong, China)*

### Abstract

Existing research has approached self-deception mainly as an intrapersonal process representing personality traits, motivational biases in information processing, and inconsistencies between explicit and implicit self-systems. Using an evolutionary approach, we treat self-deception as an interpersonal process whereby individuals deceive themselves to better deceive others. This approach advances two new theoretical developments in the understanding of self-deception. We propose that the probabilities of deception detection should affect the likelihood and activation of self-deception, and that a dual-retrieval memory system can be used to operationalize inter-personal self-deception.

We first review the existing literature by highlighting three characteristics. First, self-deceptive individuals show contradictory responses between self-report and physiological or behavioral reaction, and these inconsistencies are motivated by the desire to maintain high self-esteem. Second, self-deception is driven by self-enhancing motivations in information processing that includes perceiving, evaluating, and retrieving information. Third, self-deception is also viewed as a personality attribute, representing individual differences in viewing oneself in an overly positive light. These studies also examine affective effects of self-deception and the correlations among emotion, cognition, motivation, and decision-making behaviors.

We then present our evolutionary explanation of self-deception which we believe advances the development of this psychological construction and research both theoretically and methodologically. Theoretically, as proposed by the evolutionary biologist, Robert Trivers, self-deception is the result of the “arms race” between deception and deception detection. Because deception is prevalent in all animals including human beings, deception detection has evolved to prevent oneself from being deceived. Cues such as signs of nervousness may be leaked through changes in facial and eye muscle contraction, voice volume and pitch, and other observable physiological and bodily changes. The evolutionary adaptation to avoid these physiological cues of detection is self-deception that leaks out no cues for detection because the individual is unaware of the ongoing deception.

Methodologically, we propose a way to test the evolutionary hypothesis about self-deception. Because self-deception evolves to better escape detection, it must respond to social conditions that register the probabilities of deception detection so that individuals should self-deceive when they sense high rather than low probabilities of detection. Social status, moral standing, and the mere number of the deceived, relative to that of the deceiver, represent the detection-registering conditions that should affect the likelihood and activation of self-deception. We also propose a dual-retrieval memory procedure to operationalize self-deception. In the first retrieval that takes place in front of the deceived, the deceiver fails to retrieve encoded information due to the motivation to deceive and to escape detection. In the second retrieval after successful deception has been achieved, the deceiver retrieves the true or full information to achieve fitness gains. The difference between the two retrievals in the self-serving direction proves self-deception.

**Key words** self-deception; deception; evolutionary psychology