心 理 学 报 2008,40(6):693 ~ 700
Acta Psychologica Sinica

欺骗任务中结果评价的 FN 效应*

孙世月[1,2] 罗跃嘉[1,3]

([1]中国科学院心理研究所心理健康重点实验室,北京 100101) ([2]中国科学院研究生院,北京 100049)

([3]北京师范大学认知神经科学与学习重点实验室,北京 100875)

摘 要 在认知任务中,结果评价阶段由负性反馈信息诱发的 ERP 相对于正性反馈信息诱发的 ERP,表现出一个相对负走向的波形变化,称为反馈负波(Feedback Negativity,FN)。实验采用模拟现实生活中替人点钞情境中的欺骗行为作为任务,要求被试对看到的人民币图片按键报告其真假,同时对其中的真币图片可以做出欺骗性反应,即故意报告其为假币并力争"欺骗"计算机,从而创造出欺骗失败减钱和欺骗成功加钱两种不同效价的结果。并通过 1 元、5 元和 10 元三种不同面额的人民币图片,考察 FN 是否反映对结果信息中得失量的评价,以及不同预期强度水平对 FN 的影响。结果发现,FN 只受结果效价、而不受数额或预期强度水平的影响。此外,FN 的发生源可能位于扣带前回附近。实验结果支持了 FN 的现代二分理论,说明 FN 可能反映了基于结果与预期是否一致方面的"好"、"坏"简单快速评价。

关键词 反馈负波 FN;结果评价;现代二分理论;ERP

分类号 B842

1 引言

结果评价是人类一项重要的认知功能,指的是人们对自身行为所导致的结果或外部反馈进行评价的过程。结果评价有助于指导人们的决策,从而优化各种行为。近年来,结果评价的神经机制逐渐成为认知神经科学领域一个备受关注的研究方向。其中,研究者们利用事件相关电位(ERP)技术确定了结果评价加工中的一个重要成分反馈负波(Feedback negativity,FN,亦有称之为 Feedback related negativity,FRN 或 Feedback error – related negativity,fERN)[1~12]。FN 敏感于反馈信息的属性,是在与代表"得到"或"正性"的反馈信息所诱发的 ERP 相比较时,"失去"或"错误"的反馈刺激所诱发的 ERP 上表现出的一个相对负走向。FN 的峰潜伏期是在反馈出现后 200 ~ 300ms 内,最大波峰位于额叶中央区域(FC)的电极[5,8,9,11~13]。源定位分析结果也显示 FN 的发生源很可能位于内侧额叶,如扣带前回(Anterior Cingulated Cortex,ACC)[1,14,15],因此 FN 也被称作内侧额叶负波(Medial frontal negativity,MFN)[2]。

关于 FN 所反映的结果评价的本质,目前仍然没有定论。Miltner 等提出 FN 与错误相关负波(error – related negativity,ERN 或 error negativity,Ne)有关,认为同一种错误加工机制既产生与错误反应相关的 ERN,也产生与负性反馈相关的 FN[1]。ERN 是伴随错误反应之后的一个负走向波,FN 与 ERN 在一些方面具有相似性,如头皮分布均为内侧额叶最大、可能有同样的发生源,这也正是 FN 有时被称为 fERN 的原因。这一观点也得到了一些实验支持[16]。Gehring 等的实验利用"得钱多、少"、"失钱多、少"四种可能结果下的赌博任务分离"得失"与"相对正误"两个因素,发现在结果为"失去"时波幅比结果为"得到"时波幅更大,但在"得钱"的情况下,"得到少"这种相对错误的结果却不会诱发 FN,而两种"失钱"的情况下,"失钱少"这种相对正确的结果仍然会诱发 FN[2]。这说明 FN 对"失去"的敏感性不反映错误觉察,而是反映"得失"本身,于是对 FN 的错误加工观点提出了挑战。Holroyd 结合计算机建模与心理生理实验对此进行检验,并提出 ERN 的强化学习理论[10]。他们认为,当行为结果比期望的更糟,即出现预测误差时,负性强化学习信

收稿日期:2007 – 11 – 06

* 国家自然科学基金(30325026,30670698)、国家教育部重点项目(106025)、长江学者和创新团队发展计划资助

通讯作者:罗跃嘉,E-mail: luoyj@ bnu. edu. cn

号通过中脑多巴胺系统传至 ACC，于是产生由错误反应诱发的 ERN 与由负性反馈诱发的 FN。这一理论在生化、脑成像、临床等层面都得到了一定程度的支持。根据该理论的解释，认知系统对外部事件的评价引起的 FN 与事件的客观价值存在线性关系[13]。可以说 ERN 的强化学习理论是对 FN 的错误加工观点的改进，该理论认为，FN 不仅反映了对于错误反应的觉察或监测，更重要的是 FN 所反映的预测误差，将会调节之后的行为。

ERN 的强化学习理论认为 FN 由不期望得到的或不喜欢的结果引起，但是并没有阐明认知系统如何将一个结果评定为喜欢的或者不喜欢的。一方面，评价系统可能根据曲线函数关系来决定某一事件的喜好度，那么处于中等喜好度的结果将会引起中等波幅的 FRN，另一方面，也可能是根据一个二分函数来对喜好度做出判断，如此就不会出现中等波幅的 FRN[4]。众多的研究结果支持了后者[4,9,13]，并进一步以与情绪的动机性理论、基于行为激活与抑制的 Gray′ 理论等类似的现代二分评价理论，解释 FN 可能反映了神经系统对基于“好”、“坏”二分效价的反馈信息的早期评估[8]。这种假设也得到了脑成像方面的实验证据[17,18]。

然而，正如 Yeung 等[4]对 ERN – 强化学习理论的置疑，FN 的现代二分理论也没有解释认知系统根据什么标准将特定结果评价为“好”或“坏”。随之而引发的一个重要问题是，FN 是否反映结果与预期之间的关系。一些研究者认为 FN 可能反映了对外界事物基于预期的/非预期的这一维度而非基于奖励惩罚客观值上“好/坏”维度的快速评价加工[5,6,9,13,19,20]。例如 Holroyd 等的研究进行了两个猜测实验[20]，实验一为有得有失条件，有三种可能结果“ + 10”，“0”，“ – 10”；实验二包括两种条件，一种情况全为得钱，另一种情况全为失钱，但得失钱的可能数额均为“0”、“2. 5”和“5”。结果发现，在三种条件下，相对中等和最差的结果均比最好的结果诱发出更大的 FN，而 FN 不受绝对性客观“好/坏”结果的影响，可以推测，在各种条件下，被试均预期猜中最好的结果，与预期不一致的结果则诱发出较大的 FN。但 Hajcak 的实验利用 75% 、50% 和 25% 三种不同概率水平的可能结果考察预期的作用，以大概率下的结果作为预期条件，发现了经典的由负性反馈诱发 FN 的结果，但 FN 在预期和非预期情况下没有显著差别[8]。Cohen 最近的研究则表明“得钱”条件下对结果的预期将会影响 ERP 波幅和 EEG 时相相干与频谱能量，但“失钱”条件的 FN 却不受预期的影响[21]。总的来说，更普遍的观点认为结果评价与预期有关，一些脑成像的研究也证实这点[22 ~ 24]。

考虑到此前的大部分结果评价的研究均采用简单学习任务或者简单赌博、猜测等任务，从一定程度上限制了被试的主动卷入度，因而诸如预期之类的复杂心理过程可能得不到反映。Mai 等进行了一系列复杂认知任务中结果评价的研究[25]，其中欺骗任务要求被试对屏幕上呈现的箭头的方向作出反应，指导语中告诉被试对在三种不同的提示刺激下分别作出完全诚实、完全欺骗和自愿欺骗的反应，最后以加钱或减钱的形式反馈是否欺骗成功的信息。他们的研究结果再次证实了 FN 敏感于结果效价而不敏感于结果数额，但其中仍然没有涉及结果评价中 FN 与预期之间关系的争论。另外，现有的考察预期因素的研究一般利用可能结果出现的概率不同操纵预期因素[8,21]，以大概率的结果为预期条件，小概率的结果为非预期条件。然而正如 Folstein 等[26]提到的，根据 FN 与 N2 在潜伏期、头皮分布、发生源等方面的相似之处，FN 从本质上讲也属于 N2 家族。尤其是考虑到 FN 可能反映了结果与预期之间的不一致时，则更加难以将其与敏感于模板失匹配的前部 N2 分离开来，而且已有的大量 oddball 范式下研究结果显示敏感于失匹配的 N2 受到刺激概率的影响。因此概率因素本身是否影响 FN 目前还有待继续探讨。

针对上述问题，本研究在 Mai 等所采用的欺骗任务中结果评价的实验范式上做出改进。一方面，模拟现实生活中拾金不昧得到奖励，而“昧金”有可能不被发现便“得钱”，也有可能被发现而惩罚“失钱”的情境，使欺骗任务的实验室研究更接近于现实生活原型，提高研究的生态效度；另一方面，为了避免概率因素的混淆，本研究通过操纵可能得失钱的数额大小，激发被试不同动机水平，基于“可能得失钱的数额越大，被试的预期越强”而非“可能得失钱的概率大便存在对结果的预期，概率小便不再预期”的假设，考察不同强度的预期对 FN 是否产生影响。

2 方法

2.1 被试

17 名（男 7 女 10）来自中国农业大学的本科生，年龄 19 ~ 24 岁。所有被试均首次参加心理学实

验、身心健康、无精神神经病史、右利手、视力或矫正视力正常。

2.2　刺激材料

采用 2005 年版人民币图片作为刺激材料。图片来源:扫描仪拍摄 1 元、5 元和 10 元三种面值人民币的正面各 10 张形成数码图片,后经 Adobe Photoshop 7.0 按统一标准处理,所有图片分辨率为 72 象素/英寸,尺寸为 8cm × 4cm,视角小于 2.5°。此外,分别以 1 张 1 元、5 元和 10 元的人民币图片为素材,将图片左下角双色横号码中的数字全部修改为 0,以此标记为假币。

2.3　实验程序

实验前告诉被试这是一项检验计算机测谎软件的研究,要求被试通过将看到的真币报告为假币欺骗计算机,并力争不被发现。

通过 E - prime 在计算机上执行实验程序。被试在一间光线柔和的隔音室内,坐于一张舒适的椅子上,两眼注视屏幕中央,双眼距离屏幕 1 米。具体刺激序列如图 1 所示,首先,屏幕中央出现一个“ + ”,提醒被试集中注意力,之后出现一张人民币图片,可能为真币或假币,要求被试若看到假币,则等待随后出现“* ”时尽快按右键报告这是假币;若看到真币,则做好准备是否欺骗计算机,并在“* ”出现后 800ms 内尽快根据自己的决定做出按键反应,按左键表示报告这是真币,即诚实反应,不欺骗;按右键报告看到的是假币,即欺骗反应(左右按键在被试间进行平衡),最后将会给予被试反应结果的反馈,指导语中告诉被试因为出现假币时不可欺骗,所以若出现假币时反应出错将会给予警告;而对真币的诚实反应结果是增加相应面额的 1% 作为报酬(如 + 0.01、+ 0.05、+ 0.10),对真币的欺骗反应有两种结果:一是欺骗成功,则增加相应面额 100% 作为奖励,二是欺骗失败,则从报酬中减去相应面额的 2 倍以示惩罚。实际反馈中,在欺骗反应后,反映欺骗成功与欺骗失败的结果由计算机随机给出,各占 1/2,与被试行为反应无关。

整个实验共有 360 个 trials 出现真币和 72 个 trials 出现假币,实验结果只分析真币情况。整个实验分成 12 个小节完成,每 36 个 trials 被试自控时休息,Trials 之间间隔 1500 ~ 3000ms,正式实验持续 1 小时左右。为了排除在冒险性等方面的个体差异,从而保证叠加次数的大体一致,要求被试尽量保证三种面额条件下,做出诚实反应和欺骗反应的情况各占一半。

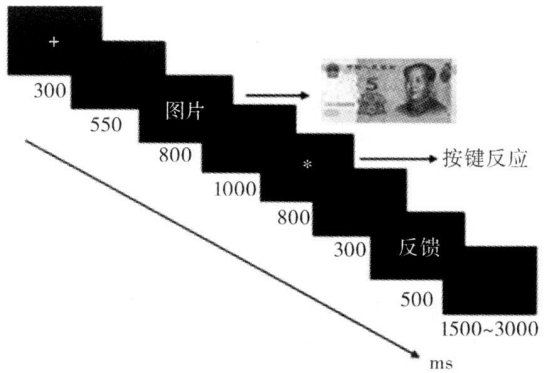

图 1　实验流程示意图

2.4　ERP 记录与分析

采用根据国际 10 - 20 系统扩展的 64 导电极帽,以 NeuroScan ERP 工作站记录 EEG 信号。头皮与电极之间阻抗小于 5 kΩ,滤波带通 0.05 - 100Hz,采样频率为每导联 500Hz。以位于左眼上下眶的电极记录垂直眼电(VEOG),位于眼外侧 1.5cm 处的左右电极记录水平眼电。以双侧乳突平均值为参考,具体是,在记录中所有电极参考置于左乳突的一只参考电极,离线分析时再次以置于右乳突的一只有效电极进行再参考,即从各导联信号中减去 1/2 该参考电极所记录的信号。

根据垂直眼电矫正眨眼伪迹,并进行 30Hz 低通滤波,自动排除其他波幅大于 ±100μV 的伪迹信号。分析时程为反馈出现前 100ms 至反馈出现后 700ms,以反馈出现前 100ms 作为基线。分别叠加并平均 1 元、5 元、10 元面额下诚实加钱、欺骗失败减钱与欺骗成功加钱三种结果反馈信息诱发的 EEG,得到根据反馈刺激锁时的 ERP 总平均波形。

采用 SPSS 13.0 统计软件对行为数据以及 ERP 波形的测量指标进行重复测量方差分析,并对不满足球形检验的统计效应采用 Greenhouse - Geisser 法进行矫正 p 值。采用 BESA 5.0 软件(Brain Electrical Source Analysis,德国 MEGIS Software GmbH 生产),对减钱条件下的 ERP 总平均波进行基于四壳椭球模型的偶极子源定位分析,以了解 FN 的发生源。

3　结果

3.1　行为结果

当人民币图片面额为 1 元、5 元、10 元时,被试选择进行欺骗的百分比分别为 47.5%、52.7%、51.9%,三者之间无显著差异。从被试进行欺骗反应和诚实反应的反应时来看,当人民币图片面额为

1 元时，分别为 323.58 ± 66.42ms 和 323.86 ± 63.24ms；当人民币图片面额为 5 元时，分别为 317.05 ± 68.22ms 和 322.80 ± 59.38ms；当人民币图片面额为 10 元时，分别为 317.21 ± 63.60ms 和 321.19 ± 62.08ms。2（反应：诚实、欺骗）× 3（数额：1、5、10）重复测量 ANOVA 结果表明各种条件下反应时没有显著差异。

3.2 ERP 波形结果

被试选择诚实和欺骗两种反应时，结果评价存在着本质区别。被试诚实反应时，就已经明确知道将要获得的结果反馈，即反应与反馈之间是完全对应的关系；而被试做出欺骗反应之后的结果存在不确定性。从 ERP 波形（图 2）也可以看出被试进行诚实反应与进行欺骗反应后反馈诱发的波形有显著差异。鉴于本研究的目的是考察结果效价对 FN 的影响，故实验结果不分析诚实反应后的确定结果，而主要关注由欺骗反应后反馈信息诱发的 ERP 波形。从图 1 可见，欺骗反应后的两种结果在反馈出现后 230 ~ 450ms 时间段内，相对于加钱（欺骗成功的情况）后的反馈结果，减钱（欺骗失败的情况）的反馈信息诱发一个负走向，这一现象即为 FN 效应。

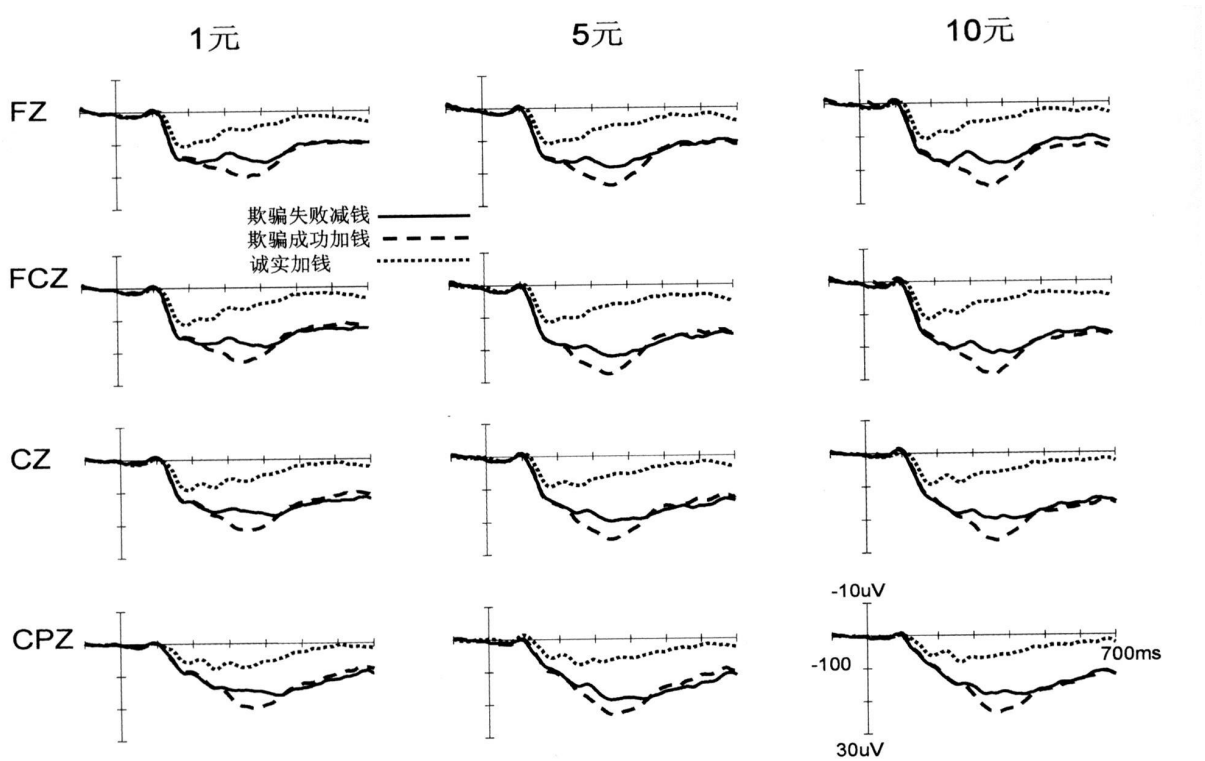

图 2 不同数额下各种反馈信息诱发的 ERP 总平均图（电极：FZ、FCZ、CZ、CPZ）

以反馈出现后 230 ~ 450ms 时间窗的平均波幅作为测量指标，进行 2（结果效价：加钱、减钱）× 3（数额：1、5、10）× 3（左右位置：左、中、右）× 6（前后位置：F 区域、FC 区域、C 区域、CP 区域、P 区域、PO 区域）四因素重复测量方差分析，有 F3、FZ、F4、FC3、FCZ、FC4、C3、CZ、C4、CP3、CPZ、CP4、P3、PZ、P4、PO3、POZ 和 PO4 共 18 个电极点进入统计分析。之所以选择该时间窗，一方面是参考相关研究[11,12,27]，其 FN 最大波峰出现于该时间窗；另一方面是因为从本研究的 ERP 总平均图看，作为相对于得钱反馈由失钱反馈所诱发的负走向主要表现在这一时间窗内。方差分析结果表明：结果效价主效应显著，$F(1,16) = 19.19$，$p < 0.001$，该时间窗内平均波幅在得钱（16.32 ± 1.05）的情况下比失钱（13.50 ± 0.81）情况下更大。数额因素主效应显著，$F(2, 32) = 15.23$，$p < 0.001$，配对比较显示其中 5 元（15.67 ± 1.04）、10 元（15.71 ± 0.89）条件下平均波幅均显著大于 1 元（13.35 ± 0.84）条件下的平均波幅，但 5 元、10 元之间差异不显著。结果效价与数额两因素间交互作用不显著。

从头皮分布上看，前后电极位置的主效应、前后位置与结果效价间的交互作用、前后位置与数额间

的交互作用均达到统计显著水平(分别为 $F(5,80)=23.74$, $\varepsilon=0.299$; $F(5,80)=11.62$, $p<0.05$, $\varepsilon=0.328$; $F(10,160)=2.36$, $p<0.05$, $\varepsilon=0.269$),表明各种条件下额中区域波幅最大,具体来讲,得钱比失钱的这种前后脑区分布差异更大,5 元、10 元条件下前后脑区分布差异比 1 元条件下大。

左右半球分布的主效应、左右分布与结果效价之间交互作用、左右分布与数额之间交互作用均达到统计显著水平: $F(2,32)=26.79$, $p<0.001$; $F(2,32)=3.53$, $p<0.05$; $F(4,64)=5.87$, $p<0.001$。表明各种条件下中线比左右两侧平均波幅更大(其中左侧为 14.70 ± 0.90,中线为 16.81 ± 0.94,右侧为 13.22 ± 0.94)。这种左右分布差异在得钱条件下比失钱条件下更大,而 5 元、10 元条件下左右分布差异比 1 元条件下更大。此外,前后与左右两个位置因素之间的显著性交互作用表明($F(10,160)=9.37$, $p<0.001$, $\varepsilon=0.432$),FCZ 电极点的波幅最大,脑区分布上中线波幅大于左右的趋势在头皮前部比后部显著。

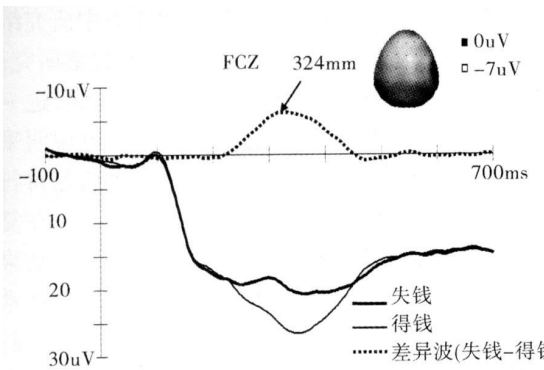

图 3　失钱和得钱的反馈诱发的 ERP 总平均波和失钱-得钱的差异波,以及差异波 324ms 时的地形图(电极:FCZ)

考虑到 FN 是相对于正性反馈的情况,负性反馈所诱发的 ERP 成分上所表现出的一个负走向,进一步以欺骗失钱反馈诱发的 ERP 减欺骗得钱反馈诱发的 ERP,得到反映 FN 的差异波(图 3)。由地形图可见,FN 的波幅在额叶中央最大。分别以 FN 的潜伏期和最大波幅为因变量,进行 3(数额:1、5、10)× 3(左右位置:左、中、右)× 6(前后位置:F 区域、FC 区域、C 区域、CP 区域、P 区域、PO 区域)的三因素重复测量方差分析(电极选择同上),结果表明,在 FN 的潜伏期上各种条件下均没有显著差异。在 FN 的波幅上,数额因素主效应不显著。在头皮分布上,左右分布主效应显著($F(2,32)=7.13$, $p<0.01$),FN 波幅在中线显著高于左右两侧(左:-7.22 ± 0.79;中:-8.33 ± 1.03;右:-7.56 ± 0.88),但左右两侧之间没有显著差异。前后分布主效应显著($F(5,80)=20.28$, $p<0.001$, $\varepsilon=0.306$),其中 FC 区波幅最大,配对比较显示 FN 波幅在额部高于中央后部。

3.3　偶极子溯源分析结果

对由减钱反馈诱发的 FN 成分进行偶极子源定位分析。为了更准确的确定 FN 成分的发生源,提高源定位的精度,以 FN 峰出现前后 50ms,即反馈信息呈现后 250 到 350ms 为时间窗做主成分分析,结果显示,单个偶极子可以解释变异的 96.1%,于是确定偶极子数为 1。不限制偶极子的方向和位置,根据最小残差标准得到偶极子定位结果(图 4)。偶极子定位在右侧扣带前回附近(Talairach 坐标系值为:x = 9.1, y = 7.2, z = 41.8),残差为 11.71%。

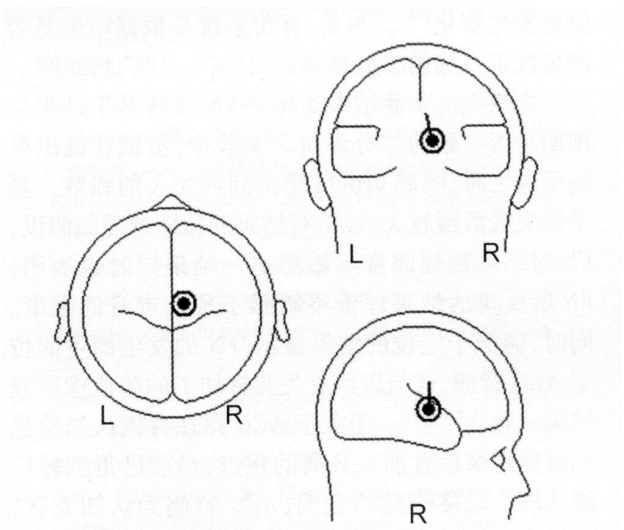

图 4　减钱反馈诱发的 FN 的偶极子源定位图(250 ~ 350ms)

4　讨论

由实验结果可见,在 230 ~ 450ms 时间窗口内,ERP 波幅受结果效价的影响,FN 更敏感于负性结果反馈信息。同时,该时间窗内结果效价与数额因素之间没有显著交互作用,以及“失钱”减“得钱”反映 FN 的差异波上数额因素主效应不显著,在一定程度上证实了 FN 对结果的数额不敏感这一观点,与当前探讨结果评价的本质的众多研究具有一致性[2,4~6,8,9,13],即 FN 只反映了评价系统将结果简单评价为“好”或“坏”的反馈加工过程,而关于结果“好”或“坏”程度的评价可能体现在其他神经活动上。

那么评价系统如何评价结果的“好/坏”呢?从

FN 对数额因素不敏感的实验结果可见，评价系统不是简单的以得失结果的绝对“好/坏”为标准的，这一结论已经由好几项研究结果给以证实[5,13,22]。于是，问题集中于结果评价与预期是否有关，不少研究已经开始了探索。尽管有一些研究发现结果评价的 FN 效应在预期与非预期两种条件下没有差异[8,21]。但更多的研究结果证实了预期因素的影响，Holroyd 等具体解释了预期如何影响结果评价[13]。他们认为，当接收到外界刺激时，认知系统首先确定当前任务目标，同时对可能的结果做出能否满足目标的二分预期，然后在结果出现之时，评价系统中的基底核根据实际结果与预期之间的差别产生预测误差，该误差信号通过中脑多巴胺系统传至扣带前回，从而产生 FN。动物实验结果也显示，在延迟反应任务中，短尾猿的眶额皮层神经活动会随着预期奖励的值而发生变化[28]。可见，评价系统是根据结果是否满足既定目标的二分预期做出“好”、“坏”判断的。

本研究的实验结果支持了 FN 反映基于结果与预期是否一致的二分评价。实验中，被试在做出欺骗反应之时，明确知道可能得到或失去的钱数。基于得失钱数额越大，被试对结果的预期越强的假设，FN 对结果数额因素不敏感这一结果同时也表明，FN 所反映的结果评价不敏感于预期本身的强度。同时，偶极子定位的结果显示，FN 的发生源可能位于 ACC 背侧，这与以往有关奖赏加工的脑成像研究结果一致[1,2,14,15]。而关于 ACC 存在背侧认知信息和腹侧情绪信息加工分离的观点，目前已得到较广泛认同。以往研究均表明，ACC 背侧为认知亚区，可能参与调节注意或执行功能、反应抑制、冲突或新异性监测、动机、工作记忆，以及认知任务的预期[29]。由此，可以认为 FN 是一个认知成分，可能反映了减钱条件下实际反馈与预期不一致所引起的冲突监测。

可以推测，可能早在结果评价之前，认知系统便对任务的目标、可能出现的结果，以及某一特定结果出现的可能性大小，特定结果是否满足目标等信息进行了综合的分析，简化为关于结果是否与预期一致的两个对立的假设。于是，当结果信息出现时，便能以快速而有效的方式完成评价过程，从而调整反应策略，优化行为结果。

此外 Krizan 等综述了有关需求偏好的研究，他们认为人类可能有一种“预期出现好结果”的乐观主义需求偏好[30]。这一假说很好的补充了 FN 的现代二分理论，例如，在本实验中，被试在进行欺骗反应之后可能存在预期自己能够欺骗成功的偏好。同时该假说也能很好的解释以往研究中所发现的一些不一致结果。例如，在 Gehring 等的实验中，首先出现标有代表得失数量“25”或“5”的卡片，在被试进行随机的猜测反应之后，通过卡片的颜色呈现“得失”的信息[2]。由于猜测的随机性，被试很可能并没有对某一结果的明确预期，只是简单的预期自己会“得钱”，于是反馈出现后，评价过程快速而直接的根据是否“得钱”将其简单的评价为“好”或“坏”，从而忽略了“失钱少”这种相对较好的结果信息。而 Hajcak 等的研究结果显示，结果的效价与概率之间未出现交互作用[8]，很可能也是被试并未根据结果可能性的大小形成预期，仍然保持了“预期出现好结果”，所以 FN 的差异只表现在两种效价上。Holroyd 等[5,20]的研究发现在确定的得失背景下，相对中等和相对最差的结果均比最好的结果诱发了更大的 FN，很可能被试总是预期出现最好的结果，更有力地支持 FN 反映了对当前结果是否与预期的好结果一致的快速二分评价。

本研究在欺骗这种复杂社会认知任务中研究结果评价，同时采用比较接近现实情境的实验室研究，得到与之前一系列研究相符的结论。一方面，进一步支持了 FN 的现代二分理论；另一方面，也说明采用这种更接近现实的欺骗任务进行结果评价是可行且有效的，关于结果评价的更深入的研究可以在复杂社会认知任务上展开，有利于更全面地了解结果评价的心理学本质。此外，通过操纵预期的强度水平，考察了预期与结果评价之间的关系，结果发现，FN 可能只反映基于“结果与既定预期是否一致”之上的“好”、“坏”二分评价，而其他结果相关信息可能发生于 FN 所反映的结果评价之前，也有可能在 FN 所反映的评价加工之外的其他神经活动进行评价。不过，本研究中受实验时间所限，各种条件下叠加次数较少（31 ± 10），反馈前间隔也较短，虽然单个被试的 ERP 结果在叠加次数最少的情况下信噪比也达到了 5.51，但不能排除实验设计在一定程度上影响了 FN 敏感性的可能，进一步的研究将会考虑这些可能的干扰因素，从而更深入的了解预期与结果评价之间的关系。此外，本研究主要从 FN 的角度解释结果评价的内在神经心理机制，着眼于结果评价本身。考虑到结果评价对行为优化的重要作用，今后研究可能进一步探索在任务完成中的动态结果评价过程。

致谢:感谢中国科学院生物物理所何士刚研究员给予本研究的宝贵建议!

参 考 文 献

1 Miltner W H R, Braun C H, Coles M G H. Event – related brain potentials following incorrect feedback in a time – estimation task: evidence for a 'generic' neural system for error – detection. Journal of Cognitive Neuroscience, 1997, 9(6): 788 ~ 798

2 Gehring W J, Willoughby A R. The medial frontal cortex and the rapid processing of monetary gains and losses. Science, 2002, 295(5563): 2279 ~ 2282

3 Nieuwenhuis S, Holroyd C B, Mol N, et al. Reinforcement – related brain potentials from medial frontal cortex: origins and functional significance. Neuroscience & Biobehavioral Reviews, 2004, 28(4): 441 ~ 448

4 Yeung N, Sanfey A G. Independent coding of reward magnitude and valence in the human brain. Journal of Neuroscience, 2004, 24(28): 6258 ~ 6264

5 Holroyd C B, Larsen J T, Cohen J D. Context dependence of the event – related brain potential associated with reward and punishment. Psychophysiology, 2004(41): 245 ~ 253

6 Toyomaki A, Murohashi H. The ERPs to feedback indicating monetary loss and gain on the game of modified "rock – paper – scissors". International Congress Series, 2005, 1278: 381 ~ 384

7 Nieuwenhuis S, Slagter H A, von Geusau N J A, et al. Knowing good from bad: differential activation of human cortical areas by positive and negative outcomes. European Journal of Neuroscience, 2005, 21(11): 3161 ~ 3168

8 Hajcak G, Holroyd C B, Moser J S, et al. Brain potentials associated with expected and unexpected good and bad outcomes. Psychophysiology, 2005, 42(2): 161 ~ 170

9 Hajcak G, Moser J S, Holroyd C B, et al. The feedback – related negativity reflects the binary evaluation of good versus bad outcomes. Biological Psychology, 2006, 71(2): 148 ~ 154

10 Holroyd C B, Coles M G H. The neuro basis of human error processing: Reinforcement learning, dopamine, and the error – related negativity. Psychological Review, 2002, 109(4): 679 ~ 709

11 Yeung N, Holroyd C B, Cohen J D. ERP Correlates of Feedback and Reward Processing in the Presence and Absence of Response Choice. Cerebral Cortex, 2005, 15(5): 535 ~ 544

12 Yu R, Zhou X. Brain potentials associated with outcome expectation and outcome evaluation. NeuroReport, 2006, 17(15): 1649 ~ 53

13 Holroyd C B, Hajcak G, Larsen J T. The good, the bad and the neutral: Electrophysiological responses to feedback stimuli. Brain Research, 2006, 1105(1): 93 ~ 101

14 Luu P, Tucker D M, Derryberry D, et al. Electrophysiological responses to errors and feedback in the process of action regulation. Psychological Science, 2003, 14(1): 47 ~ 53

15 Ruchsow M, Grothe J, Spitzer M, et al. Human anterior cingulate cortex is activated by negative feedback: evidence from event – related potentials in a guessing task. Neuroscience Letters, 2002, 325(3): 203 ~ 206

16 Nieuwenhuis S, Yeung N, Holroyd C B, et al. Sensitivity of Electrophysiological Activity from Medial Frontal Cortex to Utilitarian and Performance Feedback. Cerebral Cortex, 2004, 14(7): 741 ~ 747

17 O'Doherty J P, Buchanan T W, Seymour B, et al. Predictive neural coding of reward preference involves dissociable responses in human ventral midbrain and ventral striatum. Neuron, 2006, 49(1): 157 ~ 166

18 O'Doherty J P, Dayan P, Friston K, et al. Temporal Difference Models and Reward – Related Learning in the Human Brain. Neuron, 2003, 38(2): 329 ~ 337

19 Hajcak G, Moser J S, Holroyd C B, et al. It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. Psychophysiology, 2007, 44(6): 905 ~ 912

20 Holroyd C B, Nieuwenhuis S, Yeung N, et al. Errors in reward prediction are reflected in the event – related brain potential. Neuroreport, 2003, 14(18): 2481 ~ 2484

21 Cohen M X, Elger C E, Ranganath C. Reward expectation modulates feedback – related negativity and EEG spectra. NeuroImage, 2007, 35(2): 968 ~ 978

22 Nieuwenhuis S, Heslenfeld D J, von Geusau N J A, et al. Activity in human reward – sensitive brain areas is strongly context dependent. NeuroImage, 2005, 25(4): 1302 ~ 1309

23 Knutson B, Fong G W, Bennett S M, et al. A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event – related fMRI. Neuroimage, 2003, 18(2): 263 ~ 72

24 Cohen M X. Individual differences and the neural representations of reward expectation and reward prediction error. Social Cognitive and Affective Neuroscience, 2007, 2(1): 20 ~ 30

25 Mai X Q, Liu C, Luo Y J. Mental conflict evoked by negative feedback: An ERP Study. Progress in Biochemistry and Biophysics, 2004, 31(Suppl.): 140

26 Folstein J R, van Petten C. Influence of cognitive control and mismatch on the N2 component of the ERP: A review. Psychophysiology, 2008, 45(1): 152 ~ 170

27 Yu R, Zhou X. Brain responses to outcomes of one's own and other's performance in a gambling task. NeurpReport, 2006, 17(16): 1747 ~ 1751

28 Roesch M R, Olson C R. Neuronal activity related to reward value and motivation in primate frontal cortex. Science, 2004, 304(5668): 307 ~ 310

29 Bush G, Luu P, Posner M I. Cognitive and emotional influences in anterior cingulate cortex. Trends in Cognitive Sciences, 2000, 4(6): 215 ~ 22

30 Krizan Z, Windschitl P D. The Influence of Outcome Desirability on Optimism. Psychological Bulletin, 2007, 133(1): 95 ~ 121

Feedback-related Negativity in Outcome Evaluation with a Deception Task

SUN Shi-Yue[1,2], LUO Yue-Jia[1,3]

([1] *Key Laboratory of Mental Health, Institute of Psychology, Chinese Academy of Sciences, Beijing* 100101, *China*)

([2] *Graduate University of Chinese Academy of Sciences, Beijing* 100049, *China*)

([3] *State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing* 100875, *China*)

Abstract

Outcome evaluation is one of the important functions of the cognitive system. It can provide rapid and efficient information about the outcomes of one's behavior in order to facilitate the performance of the behavior. Recently, researchers have shown great interest in the neural mechanisms of outcome evaluation. Many studies have confirmed that a significant ERP component, called feedback-related negativity (FN), could be elicited by negative performance feedback compared to positive outcomes. The purpose of this study was to investigate the mechanisms of the outcome evaluation reflected by FN and to explore whether FN is affected by the magnitude information carrying different intensity levels of the expectation.

To observe the outcome evaluation following complex cognitive processes, a deception task was conducted in a simulated experimental situation involving the identification of currency. The participants were required to identify pictures of genuine Renminbi (RMB) from a set of pictures of fake ones. The participants were asked to press the left key to indicate genuine RMB pictures and the right key to indicate fake ones. However, we told them that for each genuine RMB picture, they could decide whether to "declare" (tell the truth) or "smuggle" (lie) and that telling the truth would result in them receiving a small but certain monetary reward, whereas lying may lead to a potential gain if they escaped being caught or a risk of double penalty if their lie was detected by the software.

Seventeen healthy undergraduates who had never participated in any electroencephalography (EEG) experiment before volunteered for this study. The EEG was recorded from 64 scalp channels using electrodes mounted in an elastic cap. Feedback-related ERPs were calculated for an 800ms epoch including a 100ms pre-feedback baseline. The outcome valences (gain when deception was successful and loss when deception was unsuccessful) by the magnitudes (pictures of the RMB worth 1, 5, and 10) resulted in six waveforms. The brain electrical source analysis (BESA) technique was also adopted in order to estimate the dipole sourcing of FN.

The ERPs of the truthful condition were obviously distinct from those of the two deceptive conditions. With regard to the deceptive conditions, compared with the "gain" feedback, the "loss" outcomes elicited a more negative deflection at the frontocentral sites in the time windows of 230 ~ 450ms. A repeated-measures ANOVA on the mean amplitudes of this time window revealed significant main effects of the outcome valences and the magnitudes; however, the interaction between these two factors did not reach significance. Further tests indicated that the "loss" outcomes elicited larger FN than did the gain outcomes and the magnitudes did not affect the FN. Finally, sourcing analysis showed that FN may be generated from brain regions near the anterior cingulated cortex (ACC).

These results suggested that the FN was sensitive to the valance rather than the magnitude of the outcome information. This finding is in agreement with the contemporary theories of outcome evaluation as well as the developed concept of the "adaptive critic" in the reinforcement learning error-related negativity (ERN) hypothesis, which suggested that FN reflected a binary evaluation of good versus bad outcomes based on whether the outcomes were consistent with the expectation.

Key words Feedback Negativity, outcome evaluation, contemporary theories of dichotomy, ERP.