The dark side of monetary incentive: how does extrinsic reward crowd out intrinsic motivation

Qingguo Ma[a,b], Jia Jin[a,b], Liang Meng[a,b] and Qiang Shen[a,b,c,d]

It was widely believed that incentives could effectively enhance the motivation of both students and employees. However, psychologists reported that extrinsic reward actually could undermine individuals' intrinsic motivation to a given interesting task, which challenged viewpoints from traditional incentive theories. Numerous studies have been carried out to test and explain the undermining effect; however, the neural basis of this effect is still elusive. Here, we carried out an electrophysiological study with a simple but interesting stopwatch task to explore to what extent the performance-based monetary reward undermines individuals' intrinsic motivation toward the task. The electrophysiological data showed that the differentiated feedback-related negativity amplitude toward intrinsic success failure divergence was prominently reduced once the extrinsic reward was imposed beforehand. However, such a difference was not observed in the control group, in which no extrinsic reward was provided throughout the experiment. Furthermore, such a pattern was not observed for P300 amplitude. Therefore, the current results indicate that extrinsic reward demotivates the intrinsic response of individuals toward success–failure outcome, which was reflected in the corresponding reduced motivational-related differentiated feedback-related negativity, but not in amplitude of P300. *NeuroReport* 25:194–198 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins.

NeuroReport 2014, **25**:194–198

Keywords: event-related potential, feedback-related negativity, performance-based reward, undermining effect

[a]School of Management, Zhejiang University, Hangzhou, [b]Neuromanagement Lab, Zhejiang University, Hangzhou, [c]National Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, China and [d]Department of Economics, Faculty of Arts and Social Sciences, National University of Singapore, Singapore, Singapore

Correspondence to Qiang Shen, PhD, Department of Economics, Faculty of Arts and Social Sciences, National University of Singapore, AS2 Level 6, 1 Arts Link, Singapore 117570, Singapore
Tel: + 65 832 28139; fax: + 86 571 879 95372; e-mail: johnsonzhj@gmail.com

Received 11 October 2013 **accepted** 27 November 2013

Introduction

In contemporary society, performance-based incentive, especially monetary reward systems, has been applied broadly to motivate employees and students in manufactories and schools, respectively. Many researchers believe that such a strategy is a practical and valid guidance of human behavior [1]. They deem that rewards can increase the likelihood that the same behavior would be repeatedly performed, which is an effect that sustains, provided that the reward contingency is operative. However, psychological studies showed an alternative kind of motivation named intrinsic reward, which was independent of external rewards, and was gained from the task *per se* [2]. Moreover, if applied improperly, the external rewards could be detrimental to the intrinsic motivation [2–4]. For example, when the monetary reward was removed, individuals' intrinsic motivation toward an interesting task would be reduced considerably compared with that in a condition without external reward.

Because of its great significance both for academic research and for practical application, the undermining effect has gained considerable attention from scientists of a broad range of disciplines including psychology and economics for decades. However, the exact mechanism for the undermining effect is still not well understood as intrinsic motivation is difficult to measure and observe directly at the behavioral level. Recently, with the rapid development of the neuroimaging techniques, a potential step toward further probing this effect is to measure brain activation at the stage of reward perception and outcome evaluation, which makes it possible for us to open the 'black box' and directly observe individuals' neural response to the intrinsic and extrinsic motivation. In a recent pioneering study, Murayama and colleagues used functional MRI to investigate the neural evidence of the motivation undermining at the spatial level. Intriguingly, they found that the BOLD signal was prominently decreased in the ventral striatum at the feedback stage when the extra reward for performance was removed at a later session of the task. Such a marked decrease in the reward valuation was not observed in the control group, where no performance-based monetary reward was provided for both sessions [5]. This suggested that the undermining effect could be reflected in the human brain's response to the success and failure of an interesting task at the feedback stage. Nevertheless, this study did not compare the dynamic changes of the neural response to the intrinsic reward valuation in a direct manner. Moreover, to the best of our knowledge, so far no study has attempted to examine the temporal dynamic mechanism of motivation undermining. Therefore, applying a between-participant design, we extended their study into a three-stage paradigm and aimed to explore the electrophysiological dynamics of the undermining effect through electroencephalography (EEG) recordings.

0959-4965 © 2014 Wolters Kluwer Health | Lippincott Williams & Wilkins DOI: 10.1097/WNR.0000000000000113

According to previous electrophysiological studies on motivation, feedback-related negativity (FRN) is the key candidate component that was found to be related to the motivation at the feedback stage in various tasks [6–8]. In Gehring and Willoughby's seminal work [9], they found a prominent differentiated FRN (d-FRN) toward the divergence of the loss gain feedback, which was considered to reflect the subjective motivational and affective evaluation of the outcome revealed. In addition, in Yeung and his colleagues' work, they also observed the FRN divergence in no active choices and no overt actions conditions, although its magnitude was reduced relative to executed ones, which further confirmed that the evaluative process indexed by FRN is sensitive to the motivational significance of an ongoing event.

To address how the external reward would impair the internal motivation and spontaneously respond toward the interesting task itself, we intended to compare electrophysiological response toward the outcome revealed before and after the imposition of extrinsic monetary reward. According to the FRN literature mentioned above, we expected that there would be a diminished FRN discrepancy after external pecuniary manipulation, reflecting the undermined subjective motivation to the intrinsic reward gained from the task.

Methods

Participants

A total of 36 healthy graduate and undergraduate students (20 men), ranging in age from 18 to 25 years (mean age = 22.30; SD = 1.83) were enrolled. They were students from Zhejiang University who did not major in psychology, business, or economics. They were all native Chinese speakers, with self-reported right-handedness. They had normal or corrected-to-normal vision, and did not have any history of neurological disorder or mental disease. The participants were assigned randomly to two groups before the event-related potential (ERP) experiment started according to the different reward scheme in the second session of the experiment. Informed consent was obtained from all participants before the commencement of the experiment and the study was approved by the Internal Review Board of Zhejiang University Neuromanagement Lab.

Stimuli

The experiment included three separate sessions for both groups. There were two blocks in each session, and each block included 45 trials, which consisted of 30 stop-watch (SW) trials and 15 watch-stop (WS) trials. The two tasks were adapted from the work by Murayama *et al.* [5]. In the SW task, a watch started automatically, and the participants were asked to stop the watch by button press and make the time fall within 70 ms deviation from the 5 s time point that was determined by a pilot study before the formal experiment, which aims to ensure that the participants can succeed in approximately half of the trials on average. In contrast, in the WS task, the participants were only asked to passively view a watch and simply press the button when it automatically stopped. The timing of the stop for a WS trial is varied between 4.2 and 5.8 s to match the time duration of SW trials in general. The trials were presented randomly in each block and the interval across trials was varied between 600 and 1000 ms. Stimuli were presented sequentially in the center of the CRT computer screen ($6.2^\circ \times 6.2^\circ$). Each trial began with a cue presented for 2000 ms indicating which task would be performed. The task started 600–1000 ms after the cue onset and outcome of the performance was revealed for 2 s and randomized a blank interval between trials that lasted 800–1200 ms.

Procedure

Participants were comfortably seated in a shield room 1 m away from a computer-controlled CRT monitor. Stimuli, recording triggers, and response were presented and recorded using E-Prime 2.0 software package (Psychology Software Tools, Pittsburgh, Pennsylvania, USA). Before the start of each session, participants were informed about the incentives of the following session. Participants from both the reward and the control groups were given a fixed 20 RMB payment immediately after they accomplished the task in sessions 1 and 3. In the second stage, however, the participants in the reward group were instructed to win monetary income on the basis of their performance, to be more specific, they would receive 1 Yuan for each successful hit during the session. The participants in the control group were incentivized in the same way as in sessions 1 and 3. Therefore, the control groups received a fixed amount of reward whereas the reward group was paid on the basis of their own performance in the second stage. The formal experiment started after a pilot practice.

Electroencephalographic recordings and analyses

The EEG was recorded (band-pass = 0.05–70 Hz; sampling rate = 500 Hz) with a Neuroscan Synamp2 Amplifier (Scan 4.3.1; Neurosoft Labs Inc., Sterling, Virginia, USA) using an electrode elastic cap with 64 Ag/AgCl electrodes according to the standard international 10–20 system. A frontal electrode site between FPz and Fz was used for ground and the left mastoid was chosen for reference. Data were transferred to the average of the left and right mastoids reference offline. Electrooculogram (EOG) was recorded from electrodes placed at 10 mm from the lateral canthi of both eyes (horizontal EOG) as well as above and below the left eye (vertical EOG). The EOG artifacts were corrected off-line for all participants during preprocessing. The experiment started only when the electrode impedances were maintained below 5 kΩ. The data were analyzed using Neuroscan 4.3.1. The EOG artifacts were corrected using

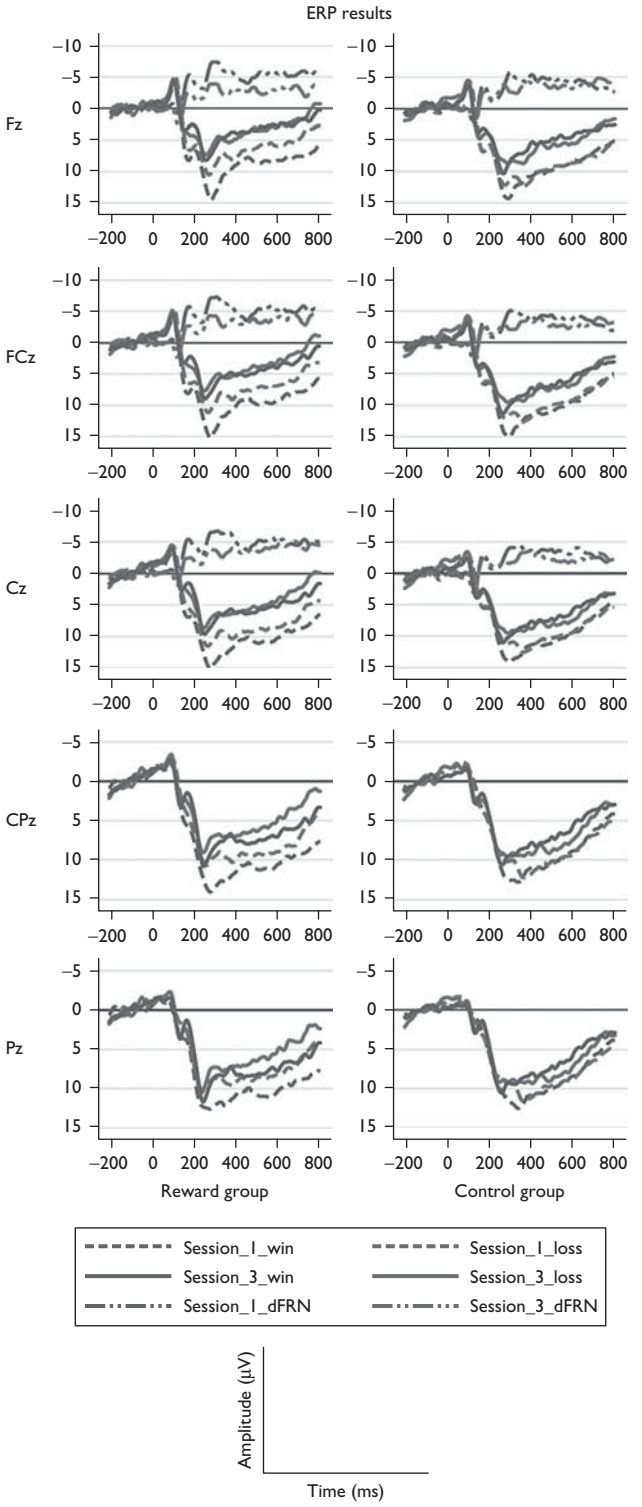
the method initially proposed by Semlitsch *et al.* [10]. Trials containing amplifier clipping, bursts of electromyography activity, or peak-to-peak deflection exceeding ±100 μV were excluded from the final analysis. ERPs were digitally filtered with a low-pass filter at 30 Hz (24 dB/octave).

The EEG recordings were segmented for the epoch from 200 ms before the onset of target to 800 ms after the onset, with the first pretargets of 200 ms as the baseline. Because of the fact that the key purpose of the current study was to investigate the influence of external reward on intrinsic motivation, we mainly focused on the ERP differences between session 1 and session 3 in the rewarded group compared with that in the control group. Therefore, in further EEG analysis, the data were collapsed on the basis of outcome of the SW task in session 1 and session 3 separately for both groups. On the basis of visual observation of grand-average waveforms and previous ERPs reports on feedback processing [11,12], two ERP components, FRN and P3, were analyzed. According to the scalp distribution of FRN and the previous studies [7,8], we chose the time range of 180–220 ms and selected nine electrode sites, namely, F1/z/2, FC1/z/2, and C1/z/2 in frontal and central areas, where it elicited the largest FRN amplitude for statistical analysis. Mixed-design analyses of variance (ANOVAs) were used to examine the effect of FRN difference of session 1 and session 3 between groups. For the analysis of P3, nine electrode sites, C1/z/2, CP1/z/2, and P1/z/2 in central and parietal areas were chosen. A similar mixed-design ANOVA was also used for P3 analysis between groups in the time window of 250–350 ms. Simple-effect analysis was carried out when there was any significant interaction effect among factors. The Greenhouse–Geisser [13] correction was applied in all statistic analyses when necessary.

Results

As shown in Fig. 1, mixed-design ANOVA results of FRN showed a significant main effect of outcome valence [$F(1,34) = 34.66$, $P < 0.001$, $\eta^2 = 0.51$], interaction effect between sessions and valence [$F(1,34) = 4.55$, $P = 0.04$, $\eta^2 = 0.12$] as well as an interaction effect among session, valence, and participant group [$F(1,34) = 4.82$, $P = 0.035$, $\eta^2 = 0.12$]. For the significant interaction effect among the three factors, simple-effect analysis was carried out in both groups separately. In the reward group, there was a significant main effect of valence [$F(1,17) = 41.69$, $P < 0.001$, $\eta^2 = 0.71$] and interaction effect over valence and session [$F(1,17) = 8.36$, $P = 0.01$, $\eta^2 = 0.33$], whereas the main effect of session was not significant [$F(1,17) < 0.1$]. Similarly, simple-effect analysis was also carried out in each session in the reward group. In session 1, a main effect of valence [$F(1,17) = 50.57$, $P < 0.001$, $\eta^2 = 0.75$] was observed. It elicited a larger FRN amplitude for loss trials than that of win trials. In the after-reward session 3, a main effect of outcome valence [$F(1,17) = 7.65$, $P = 0.01$, $\eta^2 = 0.31$]

Fig. 1



ERP results. For illustrative purpose, grand-average ERP waveforms of FRN from three frontal midline electrodes (Fz, FCz, Cz) and P300 from two parietal electrodes (Cz, CPz, Pz) were plotted as a function of session (first and third) and outcome (success and failure). In addition to this, the d-FRN was also plotted. d-FRN, differentiated feedback-related negativity; ERP, event-related potential.

was also observed. Furthermore, loss trials also induced larger FRN amplitude than win trials. For the control group, we only observed a main effect of valence [$F(1,17) = 7.41$, $P = 0.02$, $\eta^2 = 0.30$], whereas the main effect of session [$F(1,17) < 0.1$] and interaction effect [$F(1,17) < 0.1$] was not significant. In other words, the control group only showed a larger FRN amplitude toward loss trials compared with win trials, whereas the amplitude was not significantly different across sessions.

A $2 \times 2 \times 9$ mixed-design ANOVA analysis of d-FRN also showed a significant main effect for session [$F(1,34) = 4.55$, $P = 0.04$, $\eta^2 = 0.12$]; session 1 induced larger d-FRN (negative polarity: smaller voltage value means larger amplitude) than that of session 3. Interaction effect over sessions and participant groups was observed [$F(1,34) = 4.82$, $P = 0.035$, $\eta^2 = 0.12$]. Further simple-effect analysis indicated that, in the reward group, session 1 induced a significantly larger d-FRN than that of session 3 [$F(1,17) = 8.36$, $P = 0.01$, $\eta^2 = 0.33$], but there were no d-FRN differences between sessions 1 and 3 [$F(1,17) < 0.1$] in the control group.

For the analysis of P3, as also indicated in Fig. 1, there were main effects for session [$F(1,34) = 6.06$, $P = 0.02$, $\eta^2 = 0.15$], valence [$F(1,34) = 45.12$, $P < 0.001$, $\eta^2 = 0.57$], and electrode [$F(8,272) = 7.08$, $P < 0.001$, $\eta^2 = 0.17$], and there was also an interaction effect between session and valence [$F(1,34) = 10.91$, $P < 0.01$, $\eta^2 = 0.24$]. This indicated that session 1 had a larger P3 amplitude than that of session 3 and the outcome of successful hits in the SW task also induced a larger P3 amplitude than that of failed ones. Further analysis indicated a main effect of valence both in session 1 [$F(1,34) = 50.25$, $P < 0.001$, $\eta^2 = 0.60$] and in session 3 [$F(1,34) = 8.70$, $P < 0.01$, $\eta^2 = 0.20$].

Discussion

This study was carried out to explore the temporal dynamics of the undermining effect, investigating how the extrinsic monetary reward affects individuals' intrinsic motivation to a given SW task with intrinsic fun. Our data showed a prominent d-FRN discrepancy between the sessions before and after the extrinsic reward session, where the monetary incentives were awarded for good performance, whereas such a divergence was not observed in the control group, in which no extrinsic reward was provided in the midterm session. This indicates that subjective valuation toward the gain–loss outcome revealed was decreased in the third session in the reward group because of the fact that performance-based incentive was provided in the second session but was removed in the last session. This result is inconsistent with the traditional viewpoint that the external incentive can always exert a positive effect to reinforce participants' motivation to continue their work. Rather, it is in line with what we mentioned in the introduction that the extrinsic reward could, to some extent, crowd out the intrinsic motivation derived from the task *per se*.

In the current study, the incentives given in the first and third periods of the experiment were not performance based in both groups. The reduction of d-FRN in the reward group can only be attributed to the modulation effect during the second period in which participants received performance-contingent reward compared with the control group. A potential mechanism is that higher motivation leads to higher affective evaluation toward outcome information. When the intrinsic motivation was impaired, the outcome of the following task was of less affective significance, and the FRN effect decreased accordingly. This explanation is in accordance with the cognitive evaluation theory, which suggests that the undermining effect occurs because extrinsic reward ruined participants' self-determination and competency of the task [4]. When the extrinsic reward was imposed, the participants considered that they were requested to complete the task to gain a reward instead of playing for fun. Their original incentive obtained from the task itself gradually diverted to the extrinsic monetary reward. Therefore, when such an external reward was removed, they attached less importance to the outcome of task than that at the first stage, becoming less affective to the outcome at the feedback stage, resulting in the reduced amplitude of d-FRN accordingly. Moreover, recent studies also indicated that amplitude of the FRN is correlated positively with the activation of reward-related regions including the ventral striatum [14,15], which is in accordance with the recent findings of the involvement of the ventral striatum in the undermining effect using the functional MRI approach [5]. Therefore, such an observation concurs with the theory that d-FRN reflects motivational significance of the feedback outcome [9,16], which can be considered as an index of intrinsic motivation toward the given task in the current study.

Meanwhile, for the P3 component, we observed a general stage effect both for the experimental and for the control group. These findings indicate that, as the task proceeded throughout the experiment, the salience of the outcome to the participants or the attention allocation to the stimuli was reduced gradually, which is consistent with the general knowledge on P3 that it embodies the salience of the stimuli by and large [17]. In addition to this, there is a prominent effect for gain–loss discrepancy; P3 deflection loomed larger in the win condition than in the nonwin condition, which is in accordance with the recent findings that the P3 could also reflect the valence of the stimuli [6]. However, compared with the FRN, we failed to observe a prominent P3 discrepancy across groups, which might suggest that although both FRN and P3 could reflect the salience and valence of the stimuli [6,17–19], they might still play dissociated roles in the outcome evaluation process [20].

To sum up, this study investigated the neural mechanism of the undermining effect in a simple but interesting SW

task. The participants in the reward group showed reduced d-FRN divergence in the third session than that of the first session between which a performance-based monetary incentive was administered, whereas this d-FRN discrepancy effect did not appear in the control group. Our results provide evidence for the existence of the undermining effect through electrophysiological activity, which was reflected in the FRN pattern, confirming that FRN was sensitive to motivational/affective processing. This finding has empirical and theoretical significance and represents a further step toward better understanding how the interplay of extrinsic and intrinsic reward drives motivation.

Acknowledgements

This work was supported by grant 71371167 from the National Natural Science Foundation, and 211 project from Ministry of Education of China. Qiang Shen was funded by Open Research Fund of the State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University. Jia Jin was funded by grant 2012N19 from the Social Science Association of Zhejiang Province and Liang Meng was funded by grant 2013Z51 from the Social Science Association of Zhejiang Province as a key project. The authors thank Wenwei Qiu for stimuli programming.

Conflicts of interest

There are no conflicts of interest.

References

1 Skinner BF. Science and human behavior. SimonandSchuster.com; 1953.
2 Deci EL. Effects of externally mediated rewards on intrinsic motivation. *J Pers Soc Psychol* 1971; **18**:105–115.
3 Ryan RM, Mims V, Koestner R. Relation of reward contingency and interpersonal context to intrinsic motivation: a review and test using cognitive evaluation theory. *J Pers Soc Psychol* 1983; **45**:736–750.
4 Deci EL, Koestner R, Ryan RM. A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychol Bull* 1999; **125**:627–668.
5 Murayama K, Matsumoto M, Izuma K, Matsumoto K. Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proc Natl Acad Sci USA* 2010; **107**:20911–20916.
6 Wu Y, Zhou XL. The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res* 2009; **1286**:114–122.
7 Nieuwenhuis S, Yeung N, Holroyd CB, Schurger A, Cohen JD. Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex* 2004; **14**:741–747.
8 Yeung N, Sanfey AG. Independent coding of reward magnitude and valence in the human brain. *J Neurosci* 2004; **24**:6258–6264.
9 Gehring WJ, Willoughby AR. The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 2002; **295**:2279–2282.
10 Semlitsch HV, Anderer P, Schuster P, Presslich O. A solution for reliable and valid reduction of ocular artifacts, applied to the P300 ERP. *Psychophysiology* 1986; **23**:695–703.
11 Leng Y, Zhou XL. Modulation of the brain activity in outcome evaluation by interpersonal relationship: an ERP study. *Neuropsychologia* 2010; **48**:448–455.
12 Ma QG, Shen Q, Xu Q, Li DD, Shu LC, Weber B. Empathic responses to others' gains and losses: an electrophysiological investigation. *Neuroimage* 2011; **54**:2472–2480.
13 Greenhouse SW, Geisser S. On methods in the analysis of profile data. *Psychometrika* 1959; **24**:95–112.
14 Carlson JM, Foti D, Mujica-Parodi LR, Harmon-Jones E, Hajcak G. Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: a combined ERP and fMRI study. *Neuroimage* 2011; **57**:1608–1616.
15 Münte TF, Heldmann M, Hinrichs H, Marco-Pallares J, Krämer UM, Sturm V, *et al*. Nucleus accumbens is involved in human action monitoring: evidence from invasive electrophysiological recordings. *Front Hum Neurosci* 2007; **1**:11.
16 Yeung N, Holroyd CB, Cohen JD. ERP correlates of feedback and reward processing in the presence and absence of response choice. *Cerebral Cortex* 2005; **15**:535–544.
17 Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the p3, and the locus coeruleus-norepinephrine system. *Psychol Bull* 2005; **131**:510–532.
18 Bellebaum C, Polezzi D, Daum I. It is less than you expected: the feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia* 2010; **48**:3343–3350.
19 Bellebaum C, Daum I. Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *Eur J Neurosci* 2008; **27**:1823–1835.
20 Shen Q, Jin J, Ma Q. The sweet side of inequality: how advantageous status modulates empathic response to others' gains and losses. *Behav Brain Res* 2013; **256**:609–617.