doi:10.1093/scan/nst145 SCAN (2014) 9, 1530–1536

Aversive disinhibition of behavior and striatal signaling in social avoidance

Verena Ly,[1,2] Roshan Cools,[2,3,*] and Karin Roelofs[1,2,*]

[1]Radboud University Nijmegen, Behavioural Science Institute, 6500 HE Nijmegen, The Netherlands, [2]Radboud University Nijmegen, Donders Institute for Brain, Cognition, and Behaviour, Centre for Cognitive Neuroimaging, 6500 HB Nijmegen, The Netherlands and [3]Department of Psychiatry, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands

Social avoidance is a major factor contributing to the development and maintenance of anxiety and depressive symptoms. Converging evidence suggests that social avoidance is associated with abnormal aversive processing and hyperactive amygdala signaling. However, what are the consequences of such abnormal aversive processing for action and for the neural mechanisms implementing action is unclear. Existing literature is conflicting, pointing at either enhanced or reduced action inhibition. We investigated the interaction between aversion and action in social avoidance by comparing the effects of aversive *vs* appetitive faces on a go/no-go task and associated striatal signals in 42 high and low socially avoidant individuals. We combined fMRI with a novel probabilistic learning task, in which emotional valence (angry and happy faces) and optimal response (go- and no-go-responses) were manipulated independently. High compared with low socially avoidant individuals showed reduced behavioral inhibition (proportion no-go-responses) for angry relative to happy faces. This behavioral disinhibition correlated with greater striatal signal during no-go-responses for angry relative to happy faces. The results suggest that social avoidant coping style is accompanied by disinhibition of action and striatal signal in the context of social threat. The findings concur with recent theorizing about aversive disinhibition and affective disorders.

Keywords: social avoidance; aversive disinhibition; behavioral inhibition; striatum; emotional faces; go/no-go probabilistic learning task

INTRODUCTION

Social avoidance is a major risk factor for the development and maintenance of anxiety and depression (Mazer and Cloninger, 1990; Barlow, 2002; Mineka and Zinbarg, 2006). These psychiatric disorders have often been associated with enhanced aversive processing and hyperactive amygdala signaling (Schneider *et al.*, 1999; Sheline *et al.*, 2001; Veit *et al.*, 2002; Siegle *et al.*, 2006; Staugaard, 2010). Moreover, amygdala abnormalities often accompany personality traits related to social avoidance (Schwartz *et al.*, 2003; Iidaka *et al.*, 2006). Although there is converging evidence that social avoidance is associated with enhanced aversive processing, it remains unclear what are the consequences of abnormal aversive processing and aberrant amygdala signaling for action selection and for neural systems that implement action selection, such as the striatum. Insight in the emotional influence on action selection is critical for advancing our understanding of the (neurocognitive) mechanisms underlying the complex and impairing symptomatology of social avoidance.

Two competing hypotheses can be formulated on the basis of early theories and recent insights. The first and most intuitive hypothesis is that social avoidance, which is accompanied by enhanced aversive processing, is associated with an *increase* in the inhibition of actions. According to early theories on individual differences in avoidance motivation, avoidant traits and related affective disorders are associated with an enhanced tendency to respond intensely to signals of aversive stimuli, which facilitates behavioral inhibition in order to avoid punishment (Cloninger, 1987; Clark and Watson, 1991; Gray, 1994). Indeed, aversion seems to be intrinsically coupled with action inhibition: greater aversion elicits greater inhibition of actions (Boureau and Dayan, 2011). Accordingly, social avoidance, which is accompanied by enhanced aversive processing, might be associated with increased inhibition of action in an aversive context.

In contrast, based on recent neurochemical theories related to affective disorders (Dayan and Huys, 2008, 2009), one might pose the alternative, more radical hypothesis that social avoidance is accompanied by a paradoxical *decrease* in inhibition in an aversive context (i.e. aversive disinhibition). According to these latter theories, affective disorders reflect a *failure* to inhibit aversive thoughts and actions (Dayan and Huys, 2008, 2009). Individuals suffering from these affective disorders may perceive situations as more threatening and may assume more negative attributions and outcomes. In turn, this may lead to more escape and avoidance behavior as a coping strategy to enhance personal safety (Kearney, 2004). By analogy, a failure to inhibit aversive thoughts and actions might be the underlying mechanism of social avoidance. The persistent tendencies to avoid social situations in daily life might then reflect a secondary strategy to cope with this aversive disinhibition, a notion compatible with the vigilance–avoidance theory (e.g. Mogg and Bradley, 1998; Bögels and Mansell, 2004; Mogg *et al.*, 2004), and more recent interpretations of Gray's theory of the Behavioral Inhibition System (BIS; Gray and McNaughton, 2000; McNaughton and Corr, 2004; p. 286). Here we tested these two opposing hypotheses by comparing high and low socially avoidant individuals on a task that quantifies aversive inhibition.

Specifically, we used functional magnetic resonance imaging (fMRI) combined with a novel paradigm to investigate effects of aversive (relative to appetitive) processing on behavioral inhibition *vs* activation in social avoidance. Participants were presented with aversive (angry) faces and appetitive (happy) faces and had to learn by trial and error whether to make a go- or a no-go-response in order to obtain reward or avoid punishment. We manipulated emotional valence (angry/happy), the optimal response (go/no-go) and instrumental valence (reward/punishment) independently in a probabilistic learning task, in which subjects were unlikely to detect and apply an explicit rule.

Received 4 January 2013; Revised 18 June 2013; Accepted 26 August 2013
Advance Access publication 28 August 2013

We thank Paul Gaalman for assistance with data collection.

This study was supported by the Mosaic grant from the Netherlands Organization for Scientific Research (NWO) awarded to V.L. and VIDI grants from NWO awarded to K.R. and R.C. and a Human Frontiers Research Grant to R.C.

Correspondence should be addressed to Verena Ly, Radboud University Nijmegen, Donders Institute for Brain, Cognition and Behaviour, Centre for Cognitive Neuroimaging, Kapittelweg 29, 6500 HB, Nijmegen, The Netherlands. E-mail: v.ly@donders.ru.nl

*These authors contributed equally to this work

The rationale for using a probabilistic learning paradigm was to ensure that subjects recruit a (habit-based) behavioral control system that is thought to be particularly sensitive to emotional influences (Dickinson *et al.*, 1995; Holland *et al.* 2004; Dayan and Huys, 2008).

In our analyses we focused on two regions of interest. Our first region of interest was the amygdala, which is well known to be involved in aversive processing (e.g. Staugaard, 2010) and emotional accounts in motivated behavior through its interaction with other brain areas (e.g. Cardinal *et al.*, 2002). Our second region of interest was the striatum, which is known to implement action selection and to interact with the amygdala (Cardinal *et al.*, 2002). Critically, the striatum has long been shown to be associated with behavioral activation (*vs* inhibition) when facing appetitive stimuli (Schultz *et al.*, 1997; Berridge and Robinson, 1998; Salamone *et al.*, 2005; Niv *et al.*, 2007), and has recently been demonstrated to represent predominantly (go) action independent of valence (Guitart-masip *et al.*, 2011). The current paradigm was used to disentangle two sets of alternative hypotheses pointing at either aversive inhibition or aversive disinhibition underlying social avoidance. Revealing these distinct mechanisms is important, not only for understanding the neural process underlying social avoidance, but also for advancing (preventative) therapies targeted either at decreasing behavioral inhibition, or contrarily, gaining control over aversive disinhibition.

METHODS AND MATERIALS

Participants

Forty-five female students from the Radboud University Nijmegen participated in this study after giving written informed consent. We selected only women, because of the higher prevalence of affective symptoms and disorders (Kessler *et al.*, 1993; Nolen-Hoeksema, 2001) and the higher levels of emotional reactivity reported for women than men (Koch *et al.*, 2007; Domes *et al.*, 2010). They received payment or course credits as a reimbursement for participation. All participants were healthy, right-handed and had normal or corrected-to-normal visual acuity. Exclusion criteria were claustrophobia, neurological or cardiovascular diseases, psychiatric disorders, regular use of medication or marijuana, use of psychotropic drugs, heavy smoking and metal parts in the body. For one participant, the fMRI session was aborted, due to headache. Two other participants were excluded from data analyses, because their performance pattern did not meet our predefined criteria of adequate performance, which may indicate poor task compliance or motivation; in all but these two participants, simple regression analyses revealed a significant linear effect of outcome probability (i.e. the experimentally manipulated action–outcome contingencies, which determined the optimal response, significantly predicted the actual response given by the participants). Thus, data of 42 participants were analyzed. To investigate the effects of social avoidance on the interaction between emotional valence and behavioral inhibition, we divided participants in a low (low-avoidant) and a high socially avoidant (high-avoidant) group using a median-split procedure based on the avoidance subscale of the Liebowitz Social Anxiety Scale (LSAS; Liebowitz, 1987). The low-avoidant group ($M = 3.33$, $SD = 2.8$) differed significantly from the high-avoidant group ($M = 19.81$, $SD = 7.2$) in score on the avoidance subscale of the LSAS [$t(25.827) = -9.838$, $P < 0.001$].

Learning paradigm

The goal of our design was to investigate the effects of social avoidance on the influence of aversive (relative to appetitive) processing on behavioral inhibition *vs* activation, and on associated striatal BOLD signal. We manipulated emotional valence (angry/happy), optimal response (go/no-go) and instrumental valence (reward/punishment) independently in a novel probabilistic learning paradigm. Participants were presented with aversive (angry) faces and appetitive (happy) faces. They had to learn by trial and error whether to make a go- or a no-go-response (i.e. press a button, or withhold a button press, respectively) in order to obtain monetary reward or avoid monetary punishment (Figure 1). Our primary research question was focused on the effects of social avoidance on the influence of angry (relative to happy) faces on behavioral inhibition (relative to activation). We included an additional factor of instrumental outcome valence to explore the effects of social avoidance on the influence of reward and punishment anticipation on behavioral inhibition, and whether social avoidance would be associated with the extent to which reward or punishment anticipation would add to or potentiate the effects of a compatible emotional valence on behavioral inhibition.

The task consisted of four cue-types: angry-reward, angry-punishment, happy-reward and happy-punishment. We used different colors to distinguish between reward and punishment cues of the same emotional category; yellow and grayscale were randomly assigned to signal either a reward or punishment condition for each participant (which leads to either one of the two options for a participant: (i) Faces in yellow indicative of reward, and faces in grayscale indicative of punishment; or (ii) the other way around, faces in yellow indicative of punishment, and faces in grayscale indicative of reward). Participants were instructed that the combination of emotional category and color (signaling reward/punishment conditions) distinguished the four cue-types and that they had to learn the optimal response for each of the four cue-types separately. The optimal response (go- or no-go-response) was manipulated for each cue-type independently across time, by changing the action–outcome contingencies or the probability of a positive outcome given a go-response, $p(\text{Pos\_Outc}|\text{Go})$, for each cue-type over time. Specifically, for each cue-type separately, the probability of a positive outcome given a go-response could be low [$p(\text{Pos\_Outc}|\text{Go}) = 0.20$] or high [$p(\text{Pos\_Outc}|\text{Go}) = 0.80$] and was changed pseudorandomly over time. Thus, at different times during the experiment, a go-response for a certain cue-type would be rewarded or not punished on a certain percentage of trials (20% or 80%). Participants were instructed to learn the optimal response (to maximize reward, or minimize punishment) for each cue-type separately by trial and error. They were informed that the action–outcome contingencies were probabilistic and would change unpredictably over time. They were not informed about the nature of the probabilistic associations or about the time intervals across which they changed.

Each participant completed three sessions, with a 1-min break in between the sessions. Each session consisted of 160 trials, with 40 trials per cue-type. For an example of a time series, see Figure 1B. For each cue-type within a session, the probability of a positive outcome given a go-response could take one of the following combinations in two consecutive blocks: (i) 0.20, 0.20; (ii) 0.20, 0.80; (iii) 0.80, 0.80. The block lengths varied between 12 and 18 trials per cue-type, so that participants could not predict exactly when a change in contingency would occur. Moreover, to avoid instantaneous and complete reversals of the contingencies during the task, there were always short blocks (2–8 trials) of nonpredictive trials [i.e. $p(\text{Pos\_Outc}|\text{Go}) = 0.50$] at the beginning of a session and in between the blocks. We used 24 different sets of pseudorandom sequences across participants.

Timing and visual stimuli

Visual stimuli were adult Caucasian faces [trimmed to exclude influence from hair and nonfacial contours (van Peer *et al.*, 2007; Roelofs *et al.*, 2009)] from 36 models (18 men) taken from several databases (Ekman and Friesen, 1976; Matsumoto and Ekman, 1988; Martinez and Benavente, 1998; Lundqvist *et al.*, 1998). Model identity

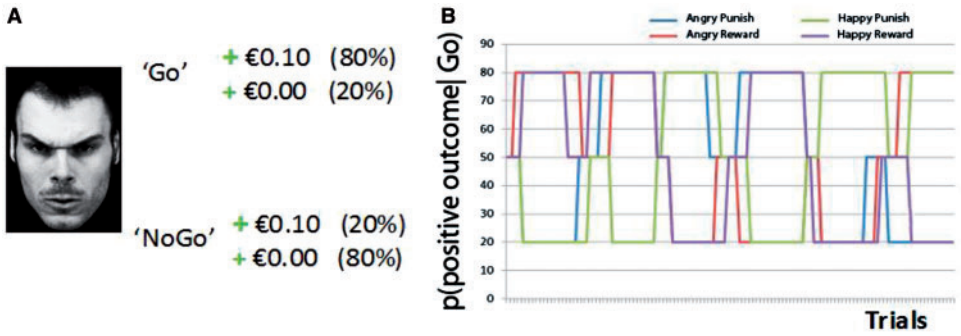

Fig. 1 Probabilistic learning paradigm. (**A**) Schematic example of a single trial. The cue was presented for 100 ms. After cue-onset, subjects were required to make a go- or no-go-response within 1000 ms. After a response–outcome delay of maximally 2000 ms, the outcome was presented (1000 ms). The duration of the intertrial interval was 3500 ms on average. In this example, the cue-type is angry-reward, the $p(\text{Pos\_Outc}|\text{Go}) = 0.80$. (**B**) Example of a pseudorandom trial series; temporal evolution of the probability of a positive outcome given a go-response, $p(\text{Pos\_Outc}|\text{Go})$ for each cue-type.

was counterbalanced, such that the model occurred equally often for each cue-type. For each model there were two emotions (angry and happy), which occurred in both yellow and grayscale (randomly assigned to signal the possibility of reward or punishment), matched for brightness and contrast values, displayed against a black background. The stimuli were projected onto a mirror above the subjects' head, subtending a visual angle of 21° by 14.6°. On each trial, one of the face cues was presented centrally for 100 ms. After cue onset, participants were required to make either a go- or a no-go-response as fast as possible within 1000 ms. If no response was made within 1000 ms, then a no-go-response was recorded. After a response–outcome delay of maximally 2000 ms (depending on the response time), the outcome was presented for 1000 ms (+10 cents for reward, −10 cents for punishment, and +0 cents for omitted reward or avoided punishment). The intertrial interval was jittered (3500 ± 1000 ms). Stimulus presentation and response acquisition were controlled by a PC running Psychophysics Toolbox Version 3 with Matlab version 7.9.0 R2009a.

Procedure

Upon arrival, the participants were reminded of the experimental procedure. They completed the LSAS. Subsequently, they were familiarized with the learning task by means of instructions (see below) and a short training session before being positioned in the MR scanner for the fMRI session to start. To increase ecological validity and participants' motivation during the learning task, we told participants that the sum of the amount of money gained and lost from the learning task would be calculated at the end of the experiment. They would receive the actual amount of monetary gain as a payment.

Behavioral data analysis

The behavioral data were analyzed using the statistics software SPSS 16.0. The proportion of no-go-responses that was made by the participants and reaction times (RTs) were analyzed using a mixed design ANOVA with group (high-avoidant/low-avoidant) as between-subject factor, and emotion (angry/happy), and outcome (instrumental valence: punishment/reward) as within-subject factors. Significant interactions were broken down by using simple interaction effects analyses. Finally, we conducted an additional analysis to test the success of the action–outcome contingency manipulation (i.e. probability of a positive outcome given a go-response). This analysis confirmed that participants were able to track the manipulation by showing the 'optimal response'.

Image acquisition

Whole-brain imaging was performed on a 3 T MR scanner (Magnetom Trio Tim; Siemens Medical Systems) equipped with an 32-channel head coil using a multi-echo GRAPPA sequence (Poser *et al.*, 2006) [repetition time (TR): 2.32 ms, echo times (TEs, 4): 9.0/19.3/30/40 ms, 38 axial oblique slices, ascending acquisition, distance factor: 17%, voxel size 3.3 × 3.3 × 2.5 mm, field of view (FoV): 211 mm; flip angle, 90°]. At the end of the experimental session, high-resolution anatomical images were acquired using a magnetization prepared rapid gradient echo sequence (TR: 2300 ms, TE: 3.03 ms, 192 sagittal slices, voxel size 1.0 × 1.0 × 1.0 mm, FoV: 256 mm).

fMRI data analysis

Images were preprocessed with SPM5, while statistical analyses were conducted with SPM8 (Statistical Parametric Mapping; http://www.fil.ion.ucl.ac.uk/spm). Given the multiecho GRAPPA MR sequence (Poser *et al.*, 2006), the head motion parameters were estimated on the MR images with the shortest TE (9.0 ms), because these images are the least affected by possible artifacts. These motion-correction parameters, estimated using a least-squares approach with six rigid body transformation parameters (translations, rotations), were then applied to the four echo images collected for each excitation. After spatial realignment, the four echo images were combined into a single MR volume using an optimized echo weighting method (Poser *et al.*, 2006). The T1-weighted image was spatially coregistered to the mean of the functional images. The fMRI time series were transformed and resampled at an isotropic voxel size of 2 mm into the standard Montreal Neurological Institute (MNI) space using both linear and nonlinear transformation parameters as determined in a probabilistic generative model that combines image registration, tissue classification, and bias correction (i.e. unified segmentation and normalization) of the coregistered T1-weighted image (Ashburner and Friston, 2005). The normalized functional images were spatially smoothed using an isotropic 8 mm full-width at half-maximum Gaussian kernel. The fMRI time series of each subject were analyzed using an event-related approach in the context of the general linear model (GLM). For each session, eight conditions of interest were modeled as separate regressors in a GLM as a function of emotion, the response that was made by the participant, and outcome: angry-punishment-go, angry-punishment-no-go, happy-punishment-go, happy-punishment-no-go, angry-reward-go, angry-reward-no-go, happy-reward-go and happy-reward-no-go. The six realignment parameters were added to capture residual head movement-related artifacts. An additional three regressors were included, describing the time course of signal intensities averaged across different image compartments of no interest (i.e. white

matter, cerebrospinal fluid and the portion of the MR image outside the skull). This procedure accounts for image intensity shifts due to movement within or near the magnetic field of the scanner (Culham *et al.*, 2003; Verhagen *et al.* 2006). All task-related regressors were modeled as delta functions at cue onset and were convolved with a canonical hemodynamic response function including time derivatives. Time series were high-pass filtered (cutoff 128 s). Temporal autocorrelation was modeled as a first-order autoregressive process.

Main (emotion, response, outcome and group) and interaction effects were analyzed using a voxel-wise threshold of $P < 0.05$ family wise error corrected for multiple comparisons across our search volumes of interest (the whole brain and small volumes of interest: the amygdala and the striatum). *A priori* hypotheses justified the selection of small volumes of interest. Specifically, we anticipated, based on previous studies (e.g. Schneider *et al.*, 1999; Veit *et al.*, 2002; Staugaard, 2010), that high-avoidant participants would show greater amygdala BOLD signals during angry *vs* happy faces compared with low-avoidant participants. The bilateral amygdala was defined anatomically using the automated anatomical labeling atlas (Tzourio-Mazoyer *et al.*, 2002). In addition, as outlined in the introduction, we were particularly interested in the effects of social avoidance on the influence of emotion on neural structures that implement behavioral activation *vs* inhibition, that is, the striatum (Guitart-masip *et al.*, 2011). To select that part of the striatum that implements behavioral activation *vs* inhibition, we adopted a functional selection procedure: the striatal ROI was defined by masking the statistical map representing the main effect of response (go/no-go) with an anatomical mask of the basal ganglia (bilateral caudate nucleus, putamen and pallidum) using the automated anatomical labeling atlas (Tzourio-Mazoyer *et al.*, 2002). To test the interaction effects, beta weights were extracted from our GLM from the individually defined amygdala and striatum ROIs and averaged over the whole ROI using MarsBaR (Brett *et al.*, 2002).

RESULTS

Behavioral results

Proportion no-go

ANOVA of the proportion of no-go-responses revealed a Group × Emotion interaction [$F(1,40) = 6.4$, $P = 0.016$], which was due to a lower proportion of no-go-responses for angry *vs* happy faces in high-avoidant participants relative to low-avoidant participants. In fact, the difference in no-go-responses between angry and happy faces was absent in high-avoidant participants [$F(1,20) = 0.203$, $P = 0.657$; Figure 2], but not in the low-avoidant participants [$F(1,20) = 18.9$, $P < 0.001$]. A main effect of outcome [$F(1,40) = 17.1$, $P < 0.001$] indicated that the proportion of no-go-responses was greater when participants avoided punishment than when they maximized reward. Finally, there was also a significant Emotion × Outcome interaction [$F(1,40) = 9.4$, $P = 0.004$] due to greater proportion of no-go-responses for angry (M = 49.8, SEM = 1.6) *vs* happy faces (M = 45.2, SEM = 1.6) in the reward condition [$F(1,40) = 9.3$, $P = 0.004$], but not in the punishment condition [$F(1,40) = 0.722$, $P = 0.4$; angry (M = 53.1, SEM = 1.3), happy (M = 54.2, SEM = 1.5)]. No other significant interaction effects were found (all $P < 0.160$). Raw data are presented in Table 1.

Reaction time on go-trials

ANOVA of RT data revealed a main effect of outcome [$F(1,40) = 10.1$, $P = 0.003$], as well as an Emotion × Outcome interaction [$F(1,40) = 5.7$, $P = 0.022$]. In addition, a main effect of emotion [$F(1,40) = 11.7$, $P = 0.001$] indicated overall faster RTs for happy *vs* angry faces. No other (interaction) effects were found (all $P > 0.434$).

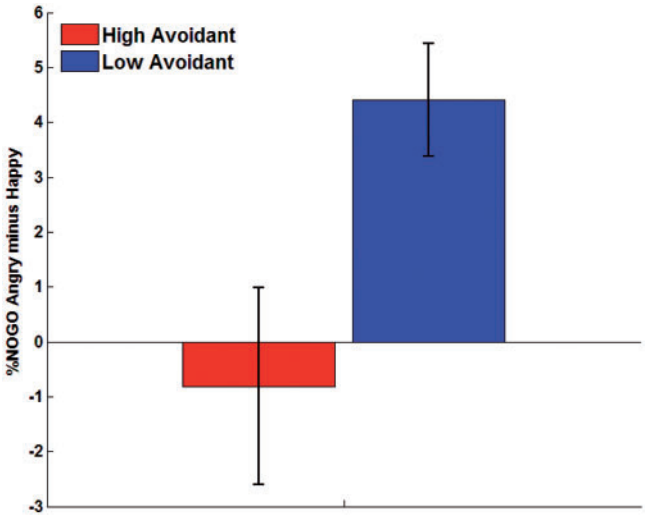

Fig. 2 Behavioral inhibition for angry minus happy faces. The *y*-axis represents the proportion of no-go-response for angry faces minus happy faces. High-avoidant individuals compared with low-avoidant individuals show significantly decreased proportion of no-go-responses for angry faces relative to happy faces, indicating decreased aversive behavioral inhibition (=aversive disinhibition). Error bars represent standard error of the difference between angry and happy faces.

Table 1 Raw data on the probabilistic learning paradigm

	Low-avoidant participants	High-avoidant participants
Angry	51.9 (1.6)	51.0 (1.8)
Happy	47.5 (1.5)	51.8 (1.9)

Proportion (%) of no-go-responses (SEM) for angry and happy faces for low-avoidant and high-avoidant groups separately.

A supplementary Emotion × Probability (of a positive outcome given a go-response) × Outcome ANOVA on the proportion of no-go-responses revealed a main effect of Probability [$F(2,39) = 135.7$, $P < 0.001$], which was due to significant differences between all probabilities in the expected direction [i.e. 20% (M = 69.3, SEM = 1.6) > 50% (M = 51.5, SEM = 1.4) > 80% (M = 31.0, SEM = 1.6); all $P < 0.001$] suggesting that participants were able to track the action–outcome contingencies.

fMRI results

Effects of emotion in the amygdala

ROI analysis of data from the anatomically defined amygdala replicated prior work (Staugaard, 2010) by showing a main effect of emotion [$F(1,40) = 7.4$, $P = 0.010$]: Amygdala signal was significantly greater for angry *vs* happy faces. Consistent with our hypothesis, we also found a Group × Emotion interaction effect [$F(1,40) = 10.9$, $P = 0.002$] due to greater amygdala response to angry *vs* happy faces in high-avoidant participants [$F(1,20) = 12.4$, $P = 0.002$] relative to low-avoidant participants [$F(1,20) = 0.316$, $P > 0.581$; Figure 3]. In addition, a main effect of response [$F(1,40) = 7.9$, $P = 0.007$] indicated that amygdala signal was greater for go- than for no-go-responses. No other significant main or interaction effects were found (all $P > 0.153$).

Effects of emotion on striatal signals associated with behavioral activation

ROI analysis of data from the functionally defined striatum confirmed a main effect of response, due to greater signals during go than

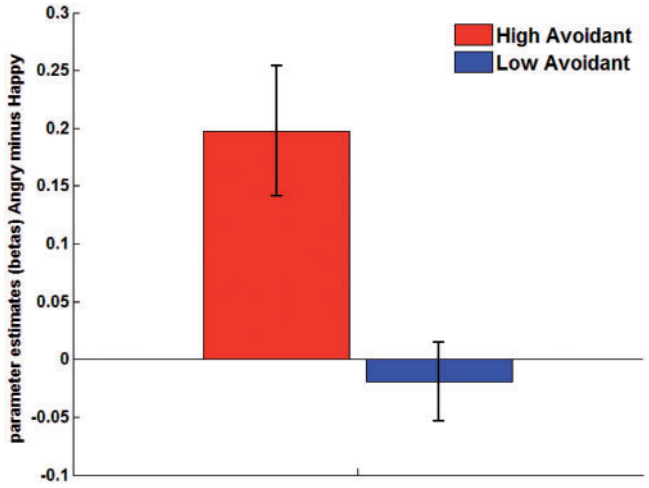

Fig. 3 Signal to face cues within anatomically defined amygdala ROI. The *y*-axis represents the mean parameter estimates as a function of Emotion. Error bars represent standard error of the difference between angry and happy faces.

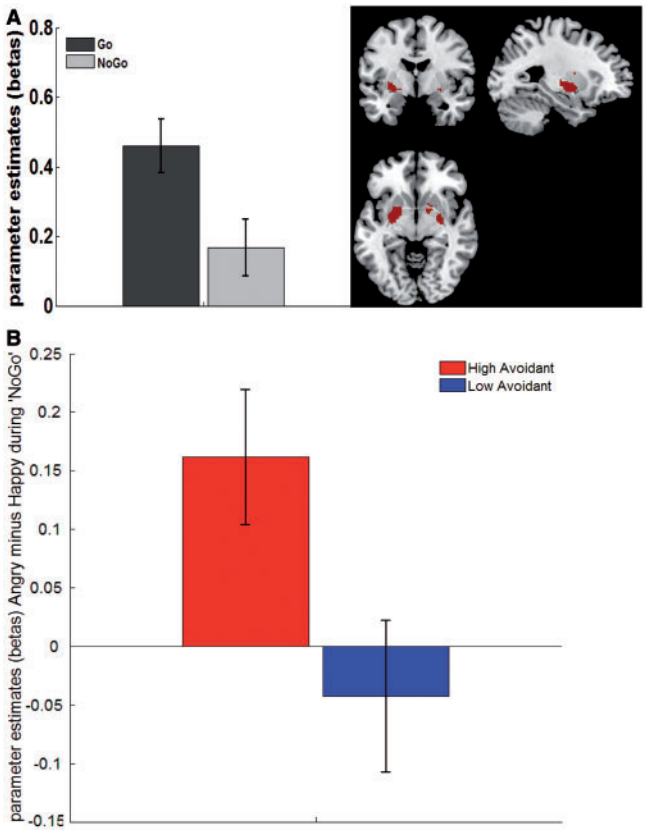

Fig. 4 Signal to face cue within functionally defined striatal region of interest. (**A**) Striatal regions of interest defined by the go *vs* no-go contrast. The left panel shows the BOLD response in this region of interest. The right panel shows the extent of the BOLD signal superimposed on a standard template obtained from software MRIcro (http://www.psychology.nottingham.ac.uk/staff/cr1/mricro.html). (**B**) Striatal BOLD signal on no-go-trials as a function of emotion and group.

no-go-trials [$F(1,40) = 67.2$, $P < 0.001$]. In addition, we found a Group × Emotion × Response interaction [$F(1,40) = 4.3$, $P < 0.044$]. To explore the nature of the Group × Emotion × Response interaction, an ANOVA was performed for high-avoidant and low-avoidant participants separately. This revealed an Emotion × Response interaction within high-avoidant participants [$F(1,40) = 7.7$, $P < 0.012$] and not within low-avoidant participants [$F(1,20) = 0.101$, $P = 0.754$]. Further analyses within the high-avoidant participants showed that the Emotion × Response interaction was due to greater striatal signal for angry *vs* happy faces in the no-go-condition [$F(1,20) = 7.8$, $P = 0.011$] and not in the go-condition [$F(1,20) = 0.607$, $P = 0.445$]. As shown in Figure 4, this enhanced striatal 'activation' signal for angry *vs* happy faces on no-go-trials was only apparent in high-avoidant participants and not in low-avoidant participants [Group × Emotion interaction; $F(1,40) = 5.5$, $P = 0.024$]. Raw data are presented in Table 2. Finally, the effect of emotion on striatal 'activation' signal during no-go-trials correlated with the effect of emotion on behavioral inhibition, as measured in terms of the proportion of no-go-responses: greater striatal 'activation' signal on angry *vs* happy no-go-trials was associated with reduced behavioral inhibition for angry *vs* happy faces ($r = -0.358$, $P = 0.020$; Figure 5). No other interaction effects were found (ROI analyses: all $P > 0.116$).

We refer to the Supplementary Materials for the results of our whole-brain voxel-wise analyses, which revealed main effects of emotion, response and outcome. No significant interaction effects were found using our stringent statistical threshold of $P_{FWE} < 0.05$.

Table 2 Mean parameter estimates extracted from the functionally defined striatal regions of interest

	Low-avoidant participants	High-avoidant participants
Angry		
Go	0.48 (0.09)	0.41 (0.12)
No-go	0.17 (0.10)	0.23 (0.14)
Happy		
Go	0.50 (0.10)	0.45 (0.14)
No-go	0.21 (0.09)	0.07 (0.15)

Data are presented as a function of emotion, and response that was made by the participant, for low-avoidant and high-avoidant groups separately (SEM).

DISCUSSION

We used fMRI combined with our novel paradigm to investigate the effects of social avoidance on the influence of aversive (relative to appetitive) processing on behavioral inhibition *vs* activation, and on associated striatal signaling. We aimed to test the two opposing hypotheses of either increased aversive inhibition or decreased aversive inhibition (aversive disinhibition) in social avoidance. Consistent with previous literature regarding emotional processing in psychiatric disorders characterized by social avoidance (e.g. Schneider *et al.*, 1999; Veit *et al.*, 2002; Staugaard, 2010), our fMRI data showed that amygdala signaling was greater for angry *vs* happy faces in high-avoidant compared with low-avoidant participants. This finding suggests enhanced aversive processing in high-avoidant compared with low-avoidant participants. However, the crucial finding of this study is that high-avoidant participants compared with low-avoidant participants showed aversive disinhibition of behavior, i.e. they showed reduced behavioral inhibition for angry faces relative to happy faces. Furthermore, this behavioral effect correlated significantly with greater striatal signaling, associated with behavioral activation, during no-go-trials for angry *vs* happy faces. Thus, our results support the hypothesis of aversive disinhibition rather than increased aversive inhibition in social avoidance.

Our findings are in line with recent neurochemical theories on behavioral inhibition in affective disorders. These theories suggest that aversive disinhibition, might be the underlying mechanism of negative thoughts and/or a lack of positive bias in psychopathology

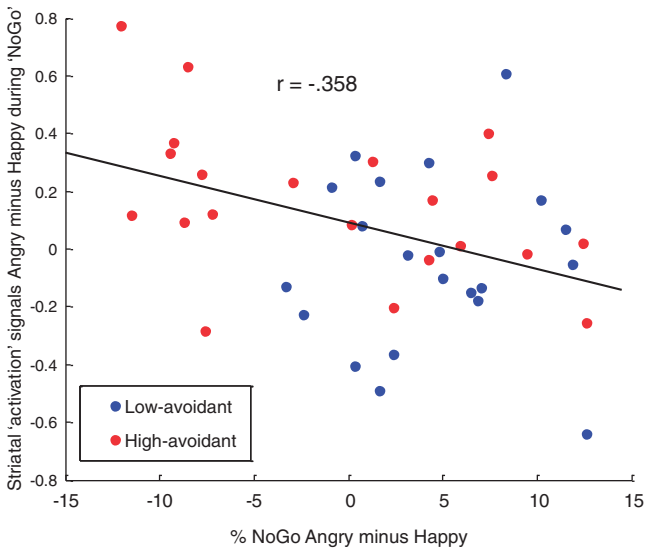

Fig. 5 Correlation between striatal 'activation' signals and behavioral inhibition. Enhanced striatal BOLD signal on no-go-trials for angry *vs* happy faces was associated with decreased proportion of no-go-responses for angry *vs* happy faces.

(Dayan and Huys, 2008, 2009). While the natural tendency to inhibit actions and thoughts leading to aversive states protects us from psychopathology, a failure to show this type of inhibition may lead to enhanced negative thoughts. Escape and avoidance behavior might then represent a coping strategy to enhance personal safety (Kearney, 2004). Accordingly, social avoidance may be a secondary consequence of aversive disinhibition. This idea is compatible with the vigilance–avoidance hypothesis, which describes the phenomenon of an enhanced initial automatic orienting to threat, followed by avoidance as a strategic attempt to alleviate the negative affective state elicited by the aversive stimuli (e.g. Mogg and Bradley, 1998; Bögels and Mansell, 2004; Mogg *et al.*, 2004). Moreover, our findings are consistent with more recent interpretations of Gray's theory of the BIS (Gray and McNaughton, 2000; McNaughton and Corr, 2004; p. 286). The BIS is proposed to play a role in resolving conflicts among competing goals by actively engaging in risk assessment behaviors. Enhanced BIS promotes scanning for threat-relevant information, and as such may act in favor for avoidance behavior.

It might be noted that the absolute behavioral data suggest less inhibition for happy faces in the low-avoidant group (Table 1). However, in the absence of a neutral control condition, we think our results cannot easily be interpreted in terms of absolute scores, but rather must be interpreted in terms of difference scores between angry relative to happy faces. Moreover, appetitive disinhibition in the low-avoidant group would be difficult to reconcile with the group difference in the fMRI data, which was driven by enhanced striatal signaling for angry relative to happy faces (during no-go-trials) in the high-avoidant participants. One could argue that this aberrant disinhibition in response to angry relative to happy faces in the high-avoidant participants could be driven by excessive inhibition to happy relative to angry faces. However, given the empirical evidence for abnormal processing of aversive faces in social avoidance in the current and previous studies (e.g. Schneider *et al.*, 1999; Veit *et al.*, 2002; Staugaard, 2010), as well as contemporary theories regarding the coupling of inhibition with aversion (Dayan and Huys, 2008, 2009), these preliminary results are best interpreted as aversive disinhibition in the high-avoidant participants. This is the first human investigation in this intriguing field of aversive disinhibition underlying socially avoidant coping and we recommend replication including neutral faces in future research to avoid interpretational limitations.

Our hypothesis that social avoidance is accompanied by abnormal influences of aversive processing, mediated by the amygdala, on action selection, mediated by the striatum, would have been strengthened by an effect of group on the functional connectivity between the amygdala and the striatum. Unfortunately, we did not detect such an effect, presumably due to low statistical power. We suggest that the open question whether aversive disinhibition in social avoidance is accompanied by abnormal amygdala–striatal connectivity should be investigated in future research.

Note that the majority of previous investigations on affective (anxiety and depressive) disorders, which show high comorbidity (Kaufman and Charney, 2000), have adopted a categorical approach, largely ignoring potentially shared mechanisms that may underlie these disorders. In line with current transdiagnostic approaches, we did not focus on social anxiety or depression per se, but we focused on a factor that has previously been described as the major maintaining factor underlying affective symptoms: social avoidance.

We developed a novel paradigm to investigate the effects of social avoidance on the emotional influence on behavioral activation. We successfully found differences between high and low socially avoidant individuals on the emotional influences on behavioral activation, as indicated by both behavioral and neural measures. Future research should examine the specificity of the findings for social avoidance *vs* its correlates (e.g. negative affectivity, mood and anxiety disorders). Finally, it remains unknown whether our behavioral and neural effects reflect the reflexive (Pavlovian) system or the more goal-driven (instrumental) system. Further research is needed, using a design which allows disentangling the Pavlovian response from the instrumental response, to investigate whether our results, suggestive of aversive disinhibition in social avoidance, involve Pavlovian control of instrumental action selection (Dayan and Huys, 2008).

In sum, we found that self-reported social avoidance tendencies are associated with aversive disinhibition of behavior and striatal signaling. These findings concur with recent behavioral neurochemical theorizing about aversion, behavioral inhibition and affective disorders and suggest that aversive disinhibition of behavior and striatal signaling might represent a core phenomenon of social avoidance behavior.

SUPPLEMENTARY DATA

Supplementary data are available at SCAN online.

Conflict of Interest

None declared.

REFERENCES

Ashburner, J., Friston, K.J. (2005). Unified segmentation. *Neuroimage*, *26*, 839–51.

Barlow, D.A. (2002). *Anxiety and its Disorders: The Nature and Treatment of Anxiety and Panic* 2nd edn. New York: Guilford Press.

Berridge, K.C., Robinson, T.E. (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*, 309–69.

Bögels, S.M., Mansell, W. (2004). Attention processes in the maintenance and treatment of social phobia: hypervigilance, avoidance and self-focused attention. *Clinical Psychological Review*, *24*, 827–56.

Boureau, Y.L., Dayan, P. (2011). Opponency revisited: competition and cooperation between dopamine and serotonin. *Neuropsychopharmacology*, *36*, 74–97.

Brett, M., Anton, J.L., Valabregue, R., Poline, J.B. (2002). Region of interest analysis using an SPM toolboxPresented at the 8th International Conference on Functional Mapping of the Human Brain, Sendai, Japan, June 2–6.

Cardinal, R.N., Parkinson, J.a., Hall, J., Everitt, B.J. (2002). Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience and Biobehavioral Reviews*, *26*, 321–52.

Clark, L.A., Watson, D. (1991). Tripartite model of anxiety and depression: psychometric evidence and taxonomic implications. *Journal of Abnormal Psychology*, *100*, 316–36.

Cloninger, C.R. (1987). A systematic method for clinical description and classification of personality variants - a proposal. *Archives of General Psychiatry*, *44*, 573–88.
Cools, R., Nakamura, K., Daw, N.D. (2010). Serotonin and dopamine: unifying affective, activational, and decision functions. *Neuropsychopharmacology*, *36*, 98–113.
Culham, J.C., Danckert, S.L., Desouza, J.F.X., Gati, J.S., Menon, R.S., Goodale, M.A. (2003). Visually guided grasping produces fMRI activation in dorsal but not ventral stream brain areas. *Experimental Brain Research*, *153*, 180–9.
Daw, N.D., Niv, Y., Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, *8*, 1701–11.
Dayan, P., Huys, Q.J.M. (2008). Serotonin, inhibition, and negative mood. *PLoS Computational Biology*, *4*, e4.
Dayan, P., Huys, Q.J.M. (2009). Serotonin in affective control. *Annual Review of Neuroscience*, *32*, 95–126.
Ekman, P., Friesen, W.V. (1976). *Pictures of Facial Affect*. Palo Alto, CA: Consulting Psychologist Press.
Dickinson, A., Balleine, B., Watt, A., Gonzalez, F., Boakes, R.A. (1995). Motivational control after extended instrumental training. *Animal Learning and Behavior*, *23*, 197–206.
Domes, G., Schultze, L., Böttger, M., et al. (2010). The neural correlates of sex differences in emotional reactivity and emotion regulation. *Human Brain Mapping*, *31*, 758–69.
Gray, J.A. (1994). Framework for a taxonomy of psychiatric disorder. In: van Goozen, S.H.M., van de Poll, N.E., Sergeant, J., editors. *Emotions: Essays on Emotion Theory*. Hillsdale, NJ: Erl baum, pp. 29–59.
Gray, J.A., McNaughton, N. (2000). *The Neuropsychology of Anxiety: An Enquiry into the Functions of the Septo-Hippocampal System*, 2nd edn. (Oxford Psychology Series No. 33). Oxford: Oxford University Press.
Guitart-masip, M., Fuentemilla, L., Bach, D.R., et al. (2011). Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. *Science*, *31*, 7867–75.
Holland, P.C. (2004). Relations between Pavlovian-instrumental transfer and reinforce devaluation. *Journal of Experimental Psychology. Animal Behavior Processes*, *30*, 104–17.
Iidaka, T., Matsumoto, A., Ozaki, N., Suzuki, T., Iwata, N. (2006). Volume of left amygdala subregion predicted temperamental trait of harm avoidance in female young subjects. A voxel-based morphometry study. *Brain Research*, *1125*, 85–93.
Kaufman, J., Charney, D. (2000). Comorbidity of mood and anxiety disorders. *Depression and Anxiety*, *12*, 69–76.
Kearney, C.A. (2004). *Social Anxiety and Social Phobia in Youth: Characteristics Assessment, and Psychological Treatment*. New York: Springer.
Kessler, R.C., McGonagle, K.A., Swartz, M., Blazer, D.G., Nelson, C.B. (1993). Sex and depression in the National Comorbidity Survey I: lifetime prevalence, chronicity, and recurrence. *Journal of Affective Disorders*, *29*, 85–96.
Koch, K., Pauly, K., Kellerman, T., et al. (2007). Gender differences in the cognitive control of emotion: an fMRI study. *Neuropsychologia*, *45*, 2744–54.
Liebowitz, M.R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–73.
Lundqvist, D., Flykt, A., Ohman, A. (1998). The Karolinska direct emotional faces (KDEF) [CD ROM]. Karolinska Institute: Sweden: Department of Clinical Neuroscience, Section of Psychology.
Martinez, A.M., Benavente, R. (1998). The AR face databaseCVC Technical Report No. 24.
Matsumoto, D., Ekman, P. (1988). Japanese and Caucasian facial expressions of emotion (JACFEE) [slides]. San Francisco (CA): University of California: Human Interaction Laboratory.
Mazer, J.D., Cloninger, R.E. (1990). *Comorbidity of mood and anxiety disorders*. Washington, DC: American Psychiatric Press.
McNaughton, N., Corr, P.J. (2004). A two-dimensional neuropsychology of defense: fear/anxiety and defensive distance. *Neuroscience and Biobehavioral Reviews*, *28*, 285–305.
Mineka, S., Zinbarg, R. (2006). A contemporary learning theory perspective on the etiology of anxiety disorders: it's not what you thought it was. *American Psychologist*, *61*(1), 10–26.
Mogg, K., Bradley, B.P. (1998). A cognitive-motivational analysis of anxiety. *Behaviour Research and Therapy*, *36*, 809–48.
Mogg, K., Bradley, B.P., Miles, F., Dixon, R. (2004). Time course of attentional bias for threat scenes: testing the vigilance-avoidance hypothesis. *Cognition and Emotion*, *18*, 689–700.
Niv, Y., Daw, N.D., Joel, D. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*, 507–20.
Nolen-Hoeksema, S. (2001). Gender differences in depression. *Current Directions in Psychological Science*, *10*, 173–6.
Poser, B.A., Versluis, M.J., Hoogduin, J.M., Norris, D.G. (2006). BOLD contrast sensitivity enhancement and artifact reduction with multiecho EPI: parallel-acquired inhomogeneity-desensitized fMRI. *Magnetic Resonance in Medicine*, *55*, 1227–35.
Roelofs, K., Minelli, A., Mars, R.B., Peer, J.V., Toni, I. (2009). On the neural control of social emotional behavior. *Social Cognitive and Affective Neuroscience*, *4*, 50–8.
Salamone, J.D., Correa, M., Mingote, S.M., Weber, S.M. (2005). Beyond the reward hypothesis: alternative functions of nucleus accumbens dopamine. *Current Opinion in Pharmacology*, *5*, 34–41.
Schneider, F., Weiss, U., Kessler, C., et al. (1999). Subcortical correlates of differential classical conditioning of aversive emotional reactions in social phobia. *Biological psychiatry*, *45*, 863–71.
Schultz, W., Dayan, P., Montague, P.R. (1997). A neural substrate of prediction and reward. *Science*, *275*, 1593–9.
Schwartz, C.E., Wright, C.I., Shin, L.M., et al. (2003). Differential amygdalar response to novel versus newly familiar neutral faces: a functional MRI probe developed for studying inhibited temperament. *Biological Psychiatry*, *53*, 854–62.
Sheline, Y.I., Barch, D.M., Donnelly, J.M., Ollinger, J.M., Snyder, A.Z., Mintun, M.A. (2001). Increased amygdala response to masked emotional faces in depressed subjects resolves with antidepressant treatment: an fMRI study. *Biological Psychiatry*, *50*, 651–8.
Siegle, G.J., Thompson, W., Carter, C.S., Steinhauer, S.R., Thase, M.E. (2006). Increased amygdala and decreased dorsolateral prefrontal BOLD responses in unipolar depression: related and independent features. *Biological Psychiatry*, *61*, 198–209.
Staugaard, S.R. (2010). Threatening faces and social anxiety: a literature review. *Clinical Psychology Review*, *30*, 669–90.
van Peer, J.M.V., Roelofs, K., Rotteveel, M., Dijk, J.G.V., Spinhoven, P., Ridderinkhof, K.R. (2007). The effects of cortisol administration on approach – avoidance behavior: an event-related potential study. *Biological Psychology*, *76*, 135–46.
Veit, R., Flor, H., Erb, M., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Interactions*, *328*, 233–36.
Verhagen, L., Grol, M.J., Dijkerman, H.C., Toni, I. (2006). Studying visuallyguided reach-to-grasp movements in an MR-environment. *Neuroimage*, *31*, S45.
Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., et al. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*, 273–89.