Supplementary Online Content

Radua J, Schmidt A, Borgwardt S, et al. Ventral striatal activation during reward processing in psychosis: a neurofunctional meta-analysis. *JAMA Psychiatry*. Published online November 11, 2015. doi:10.1001/jamapsychiatry.2015.2196

This supplementary material has been provided by the authors to give readers additional information about their work.

eMethods. Description and validation of the MetaNSUE method

First step: conversion to effect size

First, the method calculates the bounds of non statistical significance of each study with NSUEs (i.e. the unreported statistic must be within these bounds), and converts them to effect sizes using standard formula.[1] In a meta-analysis of the standardized mean difference between patients and controls, for example, the method calculates the t-values at which statistical significance would have been reached, and converts these t-values to unbiased Cohen's d. Statistics of studies with no NSUEs are also converted to effect sizes, along with all their variances.

Second step: maximum likelihood estimation (MLE)

The first step returned one effect size for each study that reported statistics, and two effect sizes corresponding to the bounds of statistical significance for each study with NSUEs, along with all their variances.

An estimation of the parameters for the subsequent imputations is then conducted by maximizing the likelihood that the reported effect-sizes have those values as well as the likelihood that the unreported effect sizes are within those bounds. Relevantly, these parameters include the between-study heterogeneity (τ^2, forced to be non-negative) and potential covariates used to individually predict the expected effect size of each study (with or without NSUEs).

Specifically, the likelihood function to maximize is the product of the probability density functions of each reported effect size and the probability mass functions that each unreported effect size lays within its two effect size bounds:

$$L\left(\theta; y_1,\ldots,y_{N_1}; y_{\alpha/2,1}, y_{1-\alpha/2,1},\ldots, y_{\alpha/2,N_2}, y_{1-\alpha/2,N_2}\right) = \prod_{i=1}^{N_1} pdf\left(y_i \mid \theta\right) \cdot \prod_{i=1}^{N_2} pmf\left(y_{\alpha/2,i}, y_{1-\alpha/2,i} \mid \theta\right)$$

where θ are the parameters (non specifically used in this text), y_i is the value of the i^{th} reported effect size, N_1 is the number of reported effect sizes, $y_{\alpha/2,i}$ $y_{1-\alpha/2,i}$ and are the effect size bounds of the i^{th} study with NSUEs, N_2 is the number of studies with NSUEs, *pdf* is the probability density function, and *pmf* is the probability mass function. Note that the latter is the difference of the cumulated distribution function evaluated at the upper and lower effect size bounds:

$$pmf\left(y_{\alpha/2}, y_{1-\alpha/2} \mid \theta\right) = cdf\left(y_{1-\alpha/2} \mid \theta\right) - cdf\left(y_{\alpha/2} \mid \theta\right)$$

where *cdf* is the distribution function.

Assuming normality, *pdf* and *cdf* are:

$$pdf\left(y_i \mid \theta\right) = \frac{1}{\sqrt{\sigma^2\left(n, y_i\right) + \tau^2}} \phi\left(\frac{y_i - X_i \cdot \beta}{\sqrt{\sigma^2\left(n, y_i\right) + \tau^2}}\right)$$

$$F\left(y_i \mid \theta\right) = \Phi\left(\frac{y_i - X_i \cdot \beta}{\sqrt{\sigma^2\left(n, y_i\right) + \tau^2}}\right)$$

where X_i is the i^{th} row of the design matrix of the predicting covariates, β is the vector of coefficients of the model of predicting covariates, $\sigma^2(n,y_i)$ is the variance of the effect size (depending on the sample size n and the reported effect size in *pdf* and the bounds of the effect size in *cdf*), ϕ is the probability density function of the standard normal distribution, and Φ is the cumulated distribution function of the standard normal distribution.

Multiplication of X and β yields the expected effect size of each study (with or without NSUEs).

Third step: multiple imputation (MI)

The second step returns the expected effect size of each study and an initial estimation of the between-study heterogeneity.

Several imputations of the unreported effect sizes are then randomly created according to the expected value, (within- and between-study) variance and statistical significance bounds of each study with NSUEs. This step is needed to create realistic "noisy" imputations, as imputing the missing effect sizes using their *expected* value would mean assuming that within-study variability and between-study heterogeneity are null.

When the variance of the effect size does not depend on the effect size (e.g. when meta-analyzing correlations) the imputations are straightforwardly created by generating random values according to the truncated normal distribution (i.e. within the bounds of non-statistical significance):

$$pdf(y_i \mid \theta) = \frac{\frac{1}{\sqrt{\sigma^2(n)+\tau^2}} \cdot \phi\left(\frac{y_i - \hat{y}}{\sqrt{\sigma^2(n)+\tau^2}}\right)}{\Phi\left(\frac{y_{1-\alpha/2} - \hat{y}}{\sqrt{\sigma^2(n)+\tau^2}}\right) - \Phi\left(\frac{y_{\alpha/2} - \hat{y}}{\sqrt{\sigma^2(n)+\tau^2}}\right)} \qquad y_i \in (y_{\alpha/2}, y_{1-\alpha/2})$$

Imputation when the variance of the effect size does depend on the effect size is less straightforward. Note that effect sizes with lower variance will receive more weight in subsequent meta-analyses, thus biasing the meta-analytic results towards them. In order to avoid this bias, the probability of a given effect size must be weighted by the inverse of the weight that the effect size would receive in a subsequent meta-analysis, with this weight being the inverse of the sum of its variance and the between-study heterogeneity:

$$pdf(y_i \mid \theta) = \frac{\left(\sigma^2(n, y_i)+\tau^2\right) \cdot \frac{1}{\sqrt{\sigma^2(n,\mu)+\tau^2}} \cdot \phi\left(\frac{y_i - \mu}{\sqrt{\sigma^2(n,\mu)+\tau^2}}\right)}{\int_{x=y_{\alpha/2}}^{y_{1-\alpha/2}} \left(\left(\sigma^2(n,x)+\tau^2\right) \cdot \frac{1}{\sqrt{\sigma^2(n,\mu)+\tau^2}} \cdot \phi\left(\frac{x - \mu}{\sqrt{\sigma^2(n,\mu)+\tau^2}}\right)\right)} \qquad y_i \in (y_{\alpha/2}, y_{1-\alpha/2})$$

Fourth step: meta-analysis and pooling

The third step returns several imputed sets of effect sizes, which are then meta-analyzed using standard formula (MetaNSUE uses restricted-maximum likelihood random-effect models for its statistical advantages.)[2] This step allows inclusion of meta-regression moderators[3]; based on simulations (see SM2 below), we recommend that they are also included in the MLE step. Finally, meta-analytic results are pooled using standard formula for MI.[4,5]

Note that most complementary analyses, such as meta-regressions or the assessment of publication bias, may be similarly conducted for each set of imputations and then pooled. However, some analyses aimed to detect influential studies by repeating the meta-analyses using different subsamples, such as the jackknife / leave-one-out, must be conducted before the MLE step in order to avoid any influence from studies not included in the subsample being analyzed.

Repeated measure analyses

Some repeated measures analyses may be straightforwardly conducted including the appropriate moderators in the meta-regression (as well as in the MLE step). This would be the case of a pre-post analysis, where the meta-regression should include one factor for time (post vs. pre) and one factor for the studies. We used this approach to compare left vs. right differences in VS activation in our meta-analysis.

Another type of repeated measures analysis, rather common in neuroimaging, is that when some samples have conducted more than one task, other samples have only conducted one task, and the meta-analytic researchers are interested in combining all the findings (but not in the differences between tasks).

One approach to deal with this situation is to combine the brain responses to the different tasks (from now on "studies") in order to have one brain response per sample.[6,7] MetaNSUE software natively includes this approach:

- During the MLE step, studies conducted with the same sample are downwards weighted so that the estimation of the parameters for the subsequent imputations is less influenced by these studies.

  Specifically, each study is weighted by:

$$w_i = \frac{1}{1 + (N_j - 1) \cdot r_{rm}}$$

  where N_j is the number of studies conducted with the same sample j as in the i^{th} study, and r_{rm} is the expected correlation between repeated measures studies (e.g. between the brain responses to two different tasks). With this adjustment, the overall weight of the studies conducted with the same sample j in the MLE step is:

$$w_j = N_j \cdot w_i = \frac{N_i}{1 + (N_j - 1) \cdot r_{rm}}$$

  which is equal to 1 if r_{rm}=1 (i.e. the different studies may be indeed considered the same), is equal to N_j if r_{rm}=0 (i.e. the N_j studies may be considered independent, not repeated

measures studies), and has an intermediate value if r_{rm}= lies between 0 and 1. These formulas may be derived, under general assumptions, from the decrease in variance of the effect size corresponding to the increase in z-value that would have been observed if the different studies conducted with the same sample *j* had been analyzed as a single study by the authors of the original studies.[7]

- During the MI step, the noise for the studies conducted with the same sample is created accounting for the correlations between them. Specifically, a linear prediction model is used to estimate the mean and variance of the normal distribution used to impute a study, conditioned to the observed and/or to the already imputed studies conducted with the same sample. The coefficients of the regression may be obtained multiplying the inverse of the variance-covariance matrix of x_1, ..., x_n with the covariance vector of x_1, ..., x_n and y, and the variance of the dependent variable subtracting the variance explained by the model from the variance of y.

- After the MI step, studies conducted with the same sample are combined in order to have a single study per sample. The effect size of the 'combined study' for the sample *j* is simply the mean of the effect sizes of the studies conducted with that sample. The variance associated to this effect size, however, is decreased by a variance reduction factor (VR_j) so that the weight of the 'combined study' is increased as above:[7]

$$VR_j = \frac{1}{w_j} = \frac{1 + (N_j - 1) \cdot r_{rm}}{N_j}$$

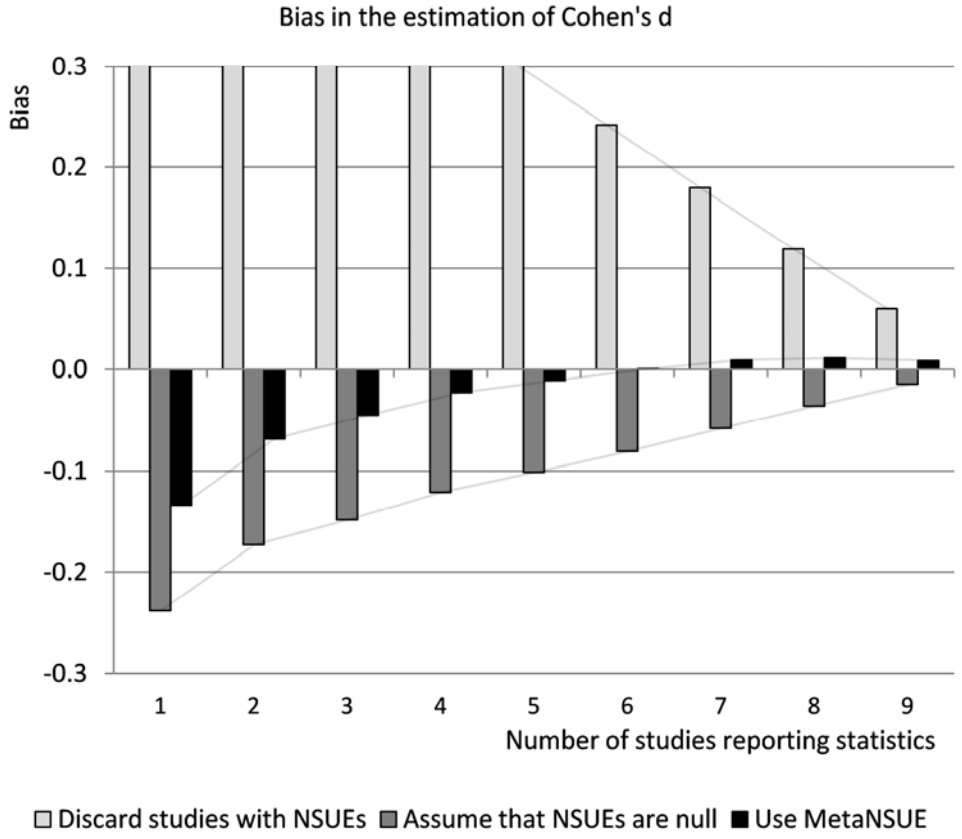
Please note that this VR_j must only be applied to the inverse of the sample sizes.[7]

Validation: comparison of MetaNSUE with standard meta-analyses

To empirically validate the MetaNSUE method, 5000 sets of 10 simulated small studies (n=20) were meta-analyzed: 1) knowing all effect-sizes (i.e. the ideal scenario); 2) discarding studies with NSUEs; 3) assuming that NSUEs are null (zero); and 4) using MetaNSUE. Results from 2-4 were then compared to 1.

As shown in the following figure, estimations where positively strongly biased when discarding studies with NSUEs (bias=0.22), moderately negatively biased when assuming that NSUEs were null (bias=-0.07), and nearly unbiased when using MetaNSUE (bias<0.001). MetaNSUE bias was moderate when only one or two studies reported effect sizes (bias=-0.08), and nearly null if at least three studies did (Figure 2). Precision was small when discarding studies with NSUEs, and

moderate otherwise (SD=0.109, 0.045 and 0.039 respectively).



Validation: inclusion vs. exclusion of moderators in the MLE

Another set of 5000 simulations was conducted to assess whether including a binary variable defining subgroups in the MLE or conducting separate meta-analyses for the two subgroups could improve or worsen the estimation of the effect size of the subgroups.

Bias when estimating subgroups and its difference was small to moderate, and higher when the binary variable was not included in the MLE (bias=0.013-0.032) than when it was included or separate analyses for the two subgroups were conducted (bias=0.008-0.019).

Validation: false positive rate when none or only one of the studies reports the effect size.

A last set of 5000 simulations was conducted to assess the false positive rate of the MetaNSUE method (i.e. simulating studies with from a population of studies with null effect size) when none or only one of the studies reported the effect size.

Global false positive rate of the MetaNSUE method was approximately 0.05 (specifically 0.038), being substantially conservative when none or only study wrongly reported a significant effect (false positive rate = 0 and 0.009 respectively). Most false positive meta-analytic results were produced when two or more studies wrongly reported a significant effect (false positive rate = 0.43), but the probability that two or more out of ten studies wrongly report a significant effect is small (0.086).

eResults

Relationship between VS activation during reward feedback and negative or positive symptoms

Five studies had investigated the relationship with negative symptoms[8-12] and we could retrieve the correlation coefficient in 1 (left) and none (right) studies, with the remaining studies reporting NSUEs. No relationship between VS activation and negative symptoms could be detected (p=0.10-1.00), though this result should be taken with caution because only five studies could be included. There was no residual heterogeneity (p=0.79-0.88). One study had investigated the relationship with positive symptoms[13] and did not report significant findings.

Relationship between VS activation during prediction error and negative or positive symptoms

Four studies had investigated the relationship with negative symptoms[14-17] and one had investigated the relationship with positive symptoms,[14] with one reporting a negative correlation between VS activation and negative symptoms,[16] one reporting the opposite relationship,[15] and two detecting no statistically-significant correlations.[14,17]

Combined meta-analysis of reward anticipation, feedback and prediction error

This combined repeated-measures meta-analysis included 40 studies (30 samples), with reported differences between patients and controls in 12 (left) and 16 (right) studies, and NSUEs in the remaining studies. Patients showed significant hypoactivation in both left and right VS (d left/right=-0.40/-0.56, $p<0.001$ in both cases, Figure 1).

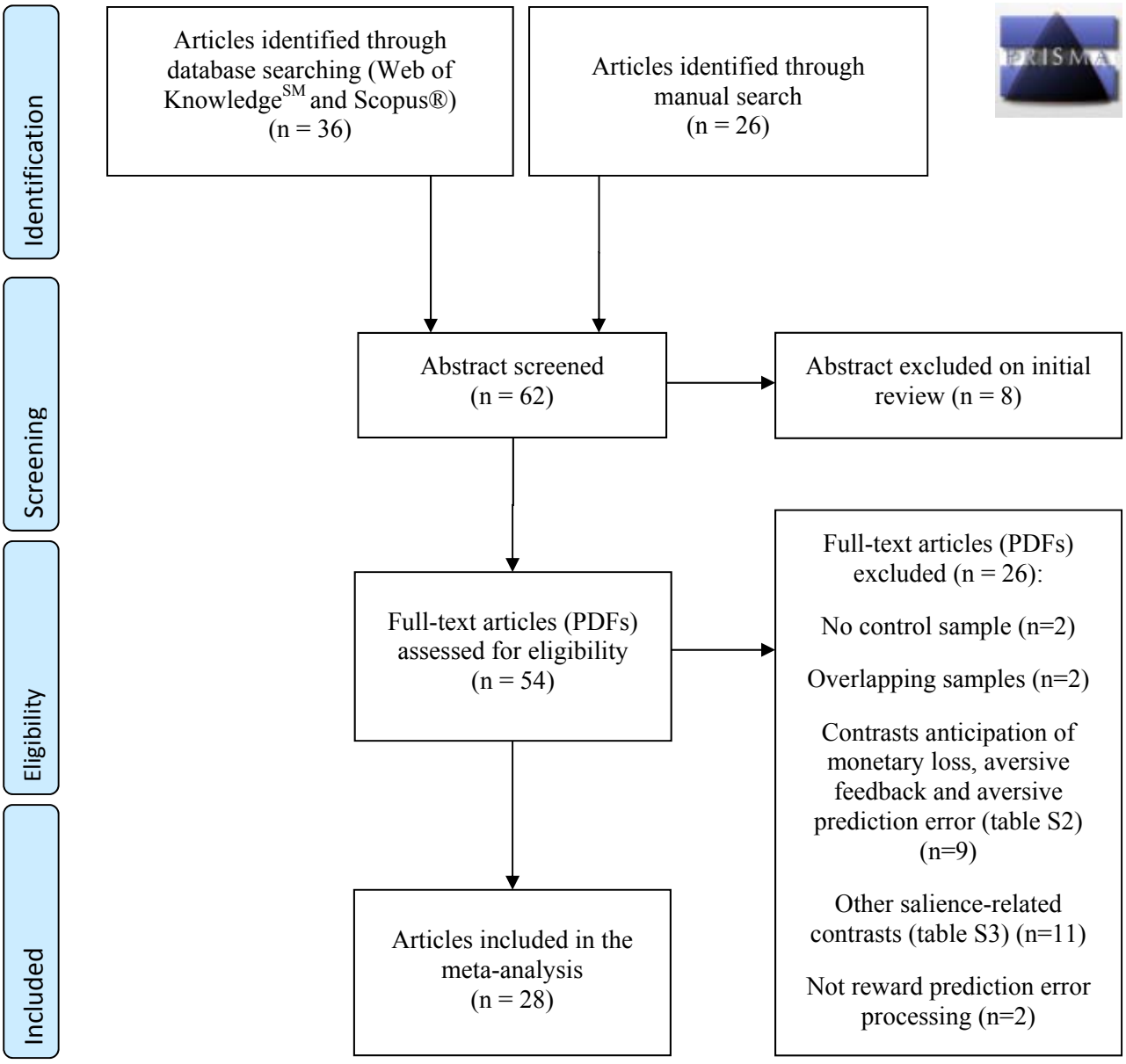
Twenty of the studies (15 samples) had analysed the correlation between VS activation in patients and negative psychotic symptoms, with reported correlation coefficients in 8 (left) and 3 (right) studies, and NSUEs in the remaining studies. At trend level ($0.008<p<0.05$), left hypoactivation was more pronounced in patients with higher scores of negative symptoms (r=-0.30, p=0.027), an effect that could not be observed at the right VS (p=0.28).

Eight of the studies (8 samples) had analysed the correlation between VS activation in patients and positive psychotic symptoms, with reported correlation coefficients in 2 (left) and 3 (right) studies, and NSUEs in the remaining studies. No relationship could be detected (p=0.40-0.88).

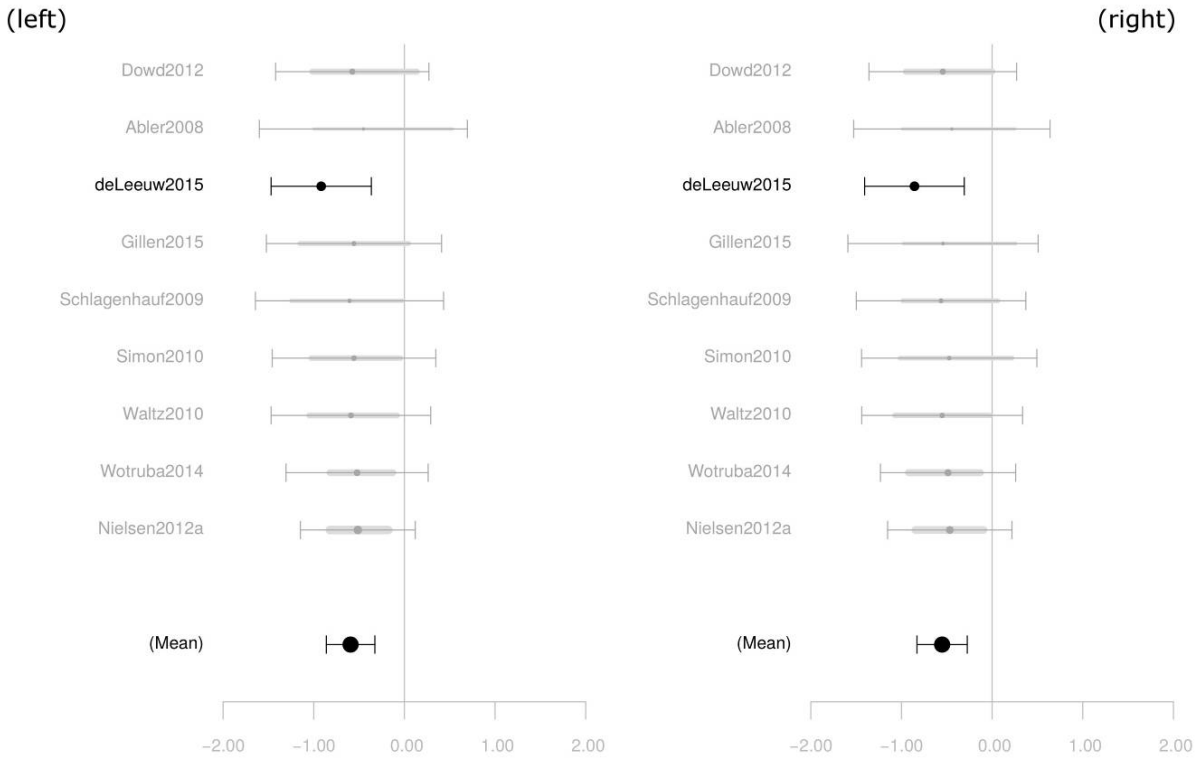
Note: Moderate but non-statistically significant residual heterogeneity was observed in all analyses (I^2: 52-74%, p=0.014-0.15). The detection of heterogeneity, however, may be underpowered for the fact that effect sizes from different domains are averaged if they were obtained in studies conducted with the same sample, homogenizing the findings. Thus, this

analysis is useful to yield a global picture, but it should be taken with caution as each domain may have its specificities.

eFigure 1. PRISMA flow chart

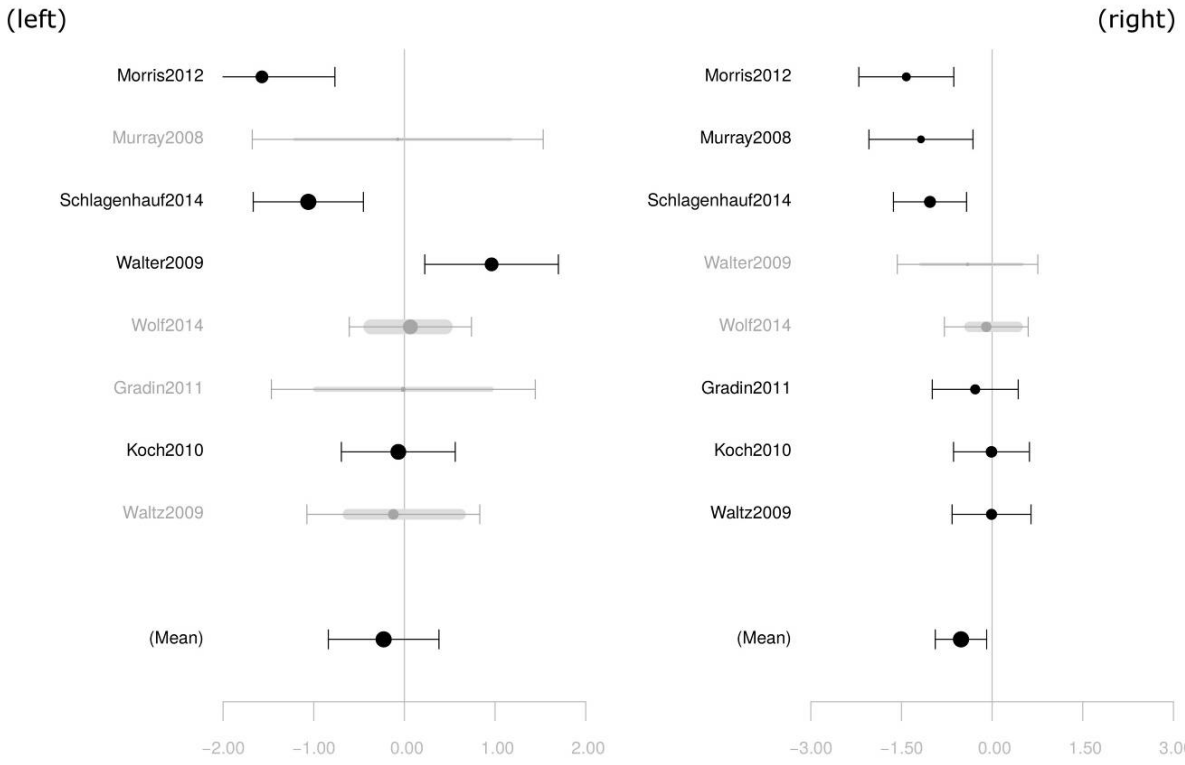


eFigure 2. Forest plots of the ventral striatum response to reward feedback in psychosis



Studies with known measures are colored black and studies with NSUEs are colored gray. For each NSUE study, the light gray shadow area shows the interval containing 95% of the imputed effect sizes, whereas the dark gray central point shows the mean.

eFigure 3. Forest plots of the ventral striatum response to reward reward prediction error in psychosis



Studies with known measures are colored black and studies with NSUEs are colored gray. For each NSUE study, the light gray shadow area shows the interval containing 95% of the imputed effect sizes, whereas the dark gray central point shows the mean.

eTable 1. MOOSE checklist

Criteria		Brief description of how the criteria were handled in the meta-analysis
Reporting of background should include		
√	Problem definition	• Contemporary theories propose that psychotic symptoms arise from abnormal processing of salient stimuli such as reward. Functional magnetic resonance imaging (fMRI) studies during reward processing tasks have detected alterations of ventral striatum activation in first-episode and chronic psychotic patients, as well as in subjects at clinical or genetic high-risk for psychosis. • However, fMRI findings seem inconclusive as some studies found reduced ventral striatum activation while other reported no abnormalities or hyperactivation. Furthermore, the relationship between symptoms and ventral striatum activation in psychosis during reward processing is unclear, with some studies revealing significant correlations with psychotic symptoms and others not.
√	Hypothesis statement	• This meta-analysis summarizes the ventral striatum activation abnormalities during monetary reward processing in psychosis, and their relationship with symptoms, using an innovative methodological approach that includes all studies even if they state that findings did not reach statistical significance but do not report statistics. • Complementary analyses (assessment of heterogeneity, potential reporting bias, meta-regressions and etcetera) are conducted to assess the robustness of the meta-analytic findings and the effects of potential clinical and methodological moderators.
√	Description of study outcomes	• Cohen's d of the differences in (right and left) ventral striatum activation (to reward anticipation, feedback of reward and reward prediction error) between subjects with ICD/DSM diagnosis of schizophrenia spectrum disorder (schizophrenia, schizophreniform, schizoaffective) or genetic and clinical high-risk state for psychosis and healthy controls. • Pearson correlation coefficient r of the relationship between (right and left) ventral striatum activation (to reward anticipation, feedback of reward and reward prediction error) in patients with psychosis and positive and negative psychotic symptoms. • The meta-analysis also includes the study of the residual heterogeneity (I^2), publication bias (meta-regression by standard error), sensitivity (Jacknife analyses), differences between left and right ventral activation (difference in Cohen's d using a repeated-measures meta-regression for side), subgroups (medication-free, patients, HR individuals, and ROI studies) and meta-regressions for age, percentage of males, percentage of medicated patients, year of publication and quality score (change in Cohen's d and p value).
√	Type of exposure or intervention used	• Functional magnetic resonance imaging (fMRI) data was acquired while patients and controls conducted monetary reward

		processing tasks (reward anticipation, feedback of reward and reward prediction error).
√	Type of study designs used	• Cross-sectional fMRI studies
√	Study population	• Subjects with ICD/DSM diagnosis of schizophrenia spectrum disorder (schizophrenia, schizophreniform, schizoaffective) or genetic and clinical high-risk state for psychosis and matched healthy controls.
Reporting of search strategy should include		
√	Qualifications of searchers	• Both investigators (AS and PFP) are PhD.
√	Search strategy, including time period included in the synthesis and keywords	• The search was extended until July 2015. The electronic research adopted several combinations of the following keywords: "psychosis", "schizophrenia", "high-risk psychosis", "salience", "fMRI", "ventral striatum", "reward", "prediction error". A second step involved the use of Scopus® and a manual search of the reference lists of the retrieved articles.
√	Databases and registries searched	• Web of Science^{SM}, MEDLINE® and Scopus®.
√	Search software used, name and version, including special features	• Web of Knowledge^{SM} and Scopus®.
√	Use of hand searching	• We hand-searched bibliographies of retrieved papers for additional references.
√	List of citations located and those excluded, including justifications	• Details of the literature search process are outlined in the Supplementary Material (PRISMA figure).
√	Method of addressing articles published in languages other than English	• The search included original articles written in English. All others were excluded.
√	Method of handling abstracts and unpublished studies	• Abstracts and unpublished studies were excluded.
√	Description of any contact with authors	• We contacted all the corresponding authors to provide additional data when needed.
Reporting of methods should include		
√	Description of relevance or appropriateness of studies assembled for assessing the hypothesis to be tested	• Detailed inclusion and exclusion criteria are described in the methods section.
√	Rationale for the selection and coding of data	• Data extracted from each of the studies were relevant to the population characteristics, study design, exposure, outcome, and possible effect of confounders.

√	Assessment of confounding	• Meta-regressions and/or subgroups were used to examine the influence of age, gender, illness stage, antipsychotic medication, use of ROIs, year of publication and quality score.
√	Assessment of study quality, including blinding of quality assessors; stratification or regression on possible predictors of study results	• We assessed the quality of the selected studies by evaluating sample size, inclusion criteria, exclusion criteria + substance abuse, match for sex / age / handedness / IQ, control for motion artifacts, co-registration with anatomical image, software and statistical test applied, ventral striatum definition, psychopathology, correction for multiple testing. Details are outlined in the Supplementary Table 2.
√	Assessment of heterogeneity	• Heterogeneity across studies was investigated with meta-regressions, and assessed calculating H^2 and the p-value of the heterogeneity test.
√	Description of statistical methods in sufficient detail to be replicated	• Description of methods of meta-analyses, meta-regression and assessment of publication bias are detailed in the methods and Supplementary Material. Open-source software is provided to straightforwardly replicate the analyses.
√	Provision of appropriate tables and graphics	• We include the PRISMA flow-chart and several tables to describe the literature search and its results. Several graphs are used to describe the main findings of the analyses.
Reporting of results should include		
√	Graph summarizing individual study estimates and overall estimate	• It is appended it in the main text.
√	Table giving descriptive information for each study included	• It is appended it in the main text.
√	Results of sensitivity testing	• Sensitivity of the findings is assessed by means of Jackknife analysis.
√	Indication of statistical uncertainty of findings	• P-values, confidence intervals and Jackknife analyses.
Reporting of discussion should include		
√	Quantitative assessment of bias	• Possibility of potential reporting bias was quantified by conducting meta-regressions by the standard error.
√	Justification for exclusion	• Exclusion criteria were: (a) other salience contrasts exploring anticipation of loss, neutral and non-monetary reward outcomes (VS findings of these studies were summarized in the Supplementary Table 3) and (b) overlapping datasets (we only included the article reporting the largest and most recent dataset).

√	Assessment of quality of included studies	• The quality assessments can be found in the supplementary Table 2. The effects of the sum score of the quality assessment are investigated using meta-regression.
Reporting of conclusions should include		
√	Consideration of alternative explanations for observed results	• We discuss potential mechanisms of our findings.
√	Generalization of the conclusions	• We include statements about the generalization of our results in the discussion and conclusions and suggest potential developments for future studies in the discussion.
√	Guidelines for future research	• They are provided in the discussion.
√	Disclosure of funding source	• All funding is declared in the main text.

eTable 2. fMRI ventral striatum findings of studies addressing anticipation of monetary loss, aversive feedback and aversive prediction error

	Patients with psychosis						***Controls***				**Ventral striatum findings (controls vs. patients)**
	Stage (a)	*N*	*Age*	*Males*	*Medicated*		*N*	*Age*	*Males*		
Anticipation of loss											
Grimm 2014[22]	HR (FGR)	54	33.6	43%	0%		80	33.5	49%		Reduced left (z = 4.61, p corrected < 0.03) and right VS activation (Z = 4.86, p corrected < 0.006) in the whole brain analysis
Hägele 2014[36]	Chronic (SZ)	44	34.2	61%	64%		54	37.7	76%		No between-group differences in VS activation
Juckel 2012[24]	HR (BS & SIPS)	13	25.5	85%	46%		13	25.7	85%		Trend for reduced left VS activation (t = 1.84, p = 0.098)
Mucci 2014[26]	Chronic (SZ)	28	33.1	64%	100%		22	31.9	45%		No between-group differences in VS activation
Silva-Alves 2013[32]	Chronic (SZ)	10	22.7	100%	100%		12	34.5	100%		Reduced left VS activation (t = 4.7, p corrected < 0.001)
Waltz 2010[35]	Chronic (SZ)	17	37.8	76%	100%		17	37.8	71%		No between-group differences in VS activation
Note: A preliminary meta-analysis of these studies showed hypoactivation in both left and right VS (d=-0.56 in both sides, p left/right=0.001/0.028). No residual heterogeneity was observed (p=0.10-0.19). In the left side, jackknife analyses showed roughly similar hypoactivation when any single study was discarded in the left side (d=[-0.39-0.67], p<0.07 in all cases). In the right side, they showed roughly similar hypoactivation when any study with NSUEs was discarded (d=[-0.62,-0.92]; p<0.015 in all cases), while a lack of activation differences was found when the only study detecting differences between patients and controls was discarded (p=0.99).										
Processing of aversive feedback											
Nielsen 2012a[10]	FEP (SZ, SA)	31	25.9	71%	0%		31	25.7	71%		Certain and uncertain loss outcome > neutral outcome: No between-group differences in VS activation

Schlagenhauf 2009[13]	Chronic (SZ)	15	30.1	80%	0%		15	30.1	80%		Successful > unsuccessful loss-avoidance: Reduced left (t = 3.38, p corrected = 0.02) and right VS activation (t = 2.77, p corrected = 0.011)
Aversive prediction error											
Romaniuk 2010[37]	Chronic (SZ, SA)	20	36.4	70%	100%		20	35.1	70%		No between-group differences in VS activation

eTable 3. fMRI ventral striatum findings of studies addressing other salience related contrasts

	Patients with psychosis						***Controls***				**Ventral striatum findings (controls vs. patients)**
	Stage (a)	*N*	*Age*	*Males*	*Medicated*		*N*	*Age*	*Males*		
Aversive Pavlovian conditioning											
Jensen 2008[38]	Chronic (SZ, SA)	13	37.6	77%	85%		13	36.5	69%		Loud noise as US: Contrast 'CS+>CS-': Reduced activation left VS (Z=3.08, p<0.05 small volume corrected) Increased right VS activation towards CS- (Z=4.14, p corrected < 0.05)
Romaniuk 2010[37]	Chronic (SZ, SA)	20	36.4	70%	100%		20	35.1	70%		IAPS pictures as US: Contrast 'CS+>CS-: No between-group differences in VS activation
Responses to emotional stimuli											
Dowd 2010[39]	Chronic (SZ, SA)	40	36.8	65%	100%		32	36.3	66%		IAPS: Reduced right VS response to positive low (F = 6.14, p < 0.03) and high arousal stimuli (F = 5.58, p < 0.03)
Grimm 2012[40]	Chronic (SZ)	23	30.3	26%	100%		23	28.9	26%		Food stimuli: Contrast 'food stimuli > control stimuli': Reduced left (T=3.99, p corrected 0.034) and right VS activation (T=3.93, p corrected 0.034)
Harvey 2010[41]	Chronic (SZ)	30	30.4	63%	90%		26	30.7	50%		No between-group differences in VS activation during positive information processing
Modinos 2015[42] (b)	FEP (P)	18	27.9	73%	56%		22	23.8	45%		No between-group differences in VS activation
	HR (UHR)	18	24.4	56%	0%						
Response to auditory oddball stimuli											

Liddle 2006[43]	Chronic (SZ, SA)	28	31.6	68%	93%		28	28.2	75%		Contrast 'target > standard stimuli': Reduced left (t=5.0, p uncorrected < 0.0005) and right VS activation (t=3.57, p uncorrected < 0.005) [reduced VS also seen for contrast 'target > novel']
Wolf 2008[44]	Chronic (SZ, SA)	17	31.9	53%	94%		21	28.6	52%		No between-group differences in VS activation
Responses during probabilistic decision-making (beads task)											
Rausch 2014[45]	Chronic (SZ)	23	33.8	70%	100%		28	35.8	54%		Reduced right VS activation (T = 2.93, p corrected = 0.038)
Rausch 2015[46]	HR (ARMS)	24	22.0	58%	0%		24	23.2	63%		Reduced right VS activation (t = 3.09, p corrected = 0.030)
Response during probabilistic reversal learning											
Waltz 2013[47]	Chronic (SZ, SA)	29	39.6	83%	100%		21	39.6	71%		No between-group differences in VS activation. Correlation with VS activation loss stay > win stay and negative symptoms SANS anhedonia & avolition

Footnote to eTables 2 and 3:

Abbreviations: ARMS, At Risk Mental State; BS, Basic Symptoms; HR, High Risk; FEP, First Episode Psychosis; FGR, First Grade Relatives of schizophrenia patients; SIPS Structured Interview for Prodromal Symptoms; P, Psychosis unspecified type; SA, SchizoAffective; SZ, Schizophrenia; UHR: Ultra High Risk.

(a) Genetic risk was defined as: i) first grade relatives of schizophrenia patients or ii) siblings of schizophrenia patients. The clinical high risk state was defined according to international and well validated criteria detailed elsewhere[48] which include: i) attenuated psychotic symptoms, ii) brief and limited intermittent psychotic symptoms, iii) genetic risk and deterioration syndrome, (iv) basic symptoms.

(b) This study (may have) included patients with diagnoses other than schizophrenia, schizoaffective or schizophreniform disorders.

eTable 4. Quality assessment of the included studies and the rating of the studies

Author & year	Sample size	Inclusion criteria	Exclusion criteria + substance abuse	Match for age/sex/ handedness/IQ	Control for motion artefacts	Co-registration with anatomical image	Software and statistical test applied	VS definition	Psychopathology	Correction for multiple testing	Sum of the scores & category
Abler 2008[18]	1	2	2	1	2	0	2	2	2	0	14
De Leeuw 2015[8]	2	2	2	2	2	2	2	2	2	2	20
Diaconescu 2011[19]	1	2	2	2	2	0	2	2	2	0	15
Dowd 2012[9]	2	2	2	1	2	2	1	2	2	2	18
Esslinger 2012[20]	2	2	1	2	2	0	2	2	2	2	17
Gillen 2015[21]	1	2	1	1	2	0	2	2	2	0	13
Gradin 2011[14]	1	2	2	1	2	0	2	2	2	2	16
Grimm 2014[22]	2	2	2	2	2	0	2	2	2	2	18
Juckel 2006[23]	0	2	2	1	2	2	2	2	2	2	17
Juckel 2012[24]	1	2	1	1	2	2	2	2	2	2	16
Koch 2010[25]	1	2	2	1	2	0	2	2	2	0	14
Morris 2012[15]	1	2	1	1	2	2	2	2	2	2	17
Mucci 2014[26]	2	2	2	1	2	0	1	2	2	2	16
Murray 2008[27]	1	2	1	2	0	0	2	2	0	2	12
Nielsen 2012a[10]	2	2	2	1	2	2	2	2	2	2	19
Nielsen 2012b[28]	2	2	2	1	2	2	2	2	2	0	17
Roiser 2013[29]	1	2	1	2	2	0	2	2	2	0	14
Schlagenhauf 2008[30]	0	2	2	2	2	2	2	2	2	2	18
Schlagenhauf 2009[13]	1	2	2	2	2	2	2	2	2	2	19
Schlagenhauf 2014[31]	2	2	2	2	2	2	2	2	2	2	20
Silva Alves 2013[32]	0.5	2	2	1	2	2	2	0	2	2	15.5
Simon 2010[11]	1	2	2	2	2	2	2	2	2	0	17
Smieskova 2015[33]	1.5	2	2	2	2	0	2	2	2	2	17.5
Walter 2009[34]	1	2	2	1	2	0	1	0	2	2	13
Waltz 2009[16]	1	2	2	2	2	2	2	2	2	2	19
Waltz 2010[35]	1	2	2	1	2	2	2	2	2	2	18
Wolf 2014[17]	2	2	2	1	2	2	2	2	2	0	17

Wotruba 2014[12]	2	2	2	2	2	2	1	2	2	2	19

Rating criteria: Sample size: n1<12,n2<12: 0 point; n1<12,n2=12-20: 0.5 point; n1<12,n2>20: 1 point; n1=12-20,n2<12: 0.5 point; n1=12-20,n2=12-20: 1 point; n1=12-20,n2>20: 1.5 point; n1>20,n2<12: 1 point; n1>20,n2=12-20: 1.5 point; n1>20,n2>20: 2 point; Inclusion criteria: 0 (not reported), 1 (partly reported), 2 (reported); Exclusion criteria + substance abuse: 0(not reported), 1 (only one reported), 2 (reported); Matched for age/sex/handedness/IQ: 0 (for no parameter), 1 (partly), 2 (for all parameters); control for motion artefacts: 0 (not performed), 2 (performed); Co-registration with anatomical image: 0 (not performed), 2 (performed); Software and statistical test applied: 0 (not reported), 1 (partly reported), 2 (reported); VS definition: 0 (not described), 2 (described, including whole brain approaches); Psychopathology: 0 (insufficiently or not reported), 2 (for all groups reported); Correction for multiple testing: 0 (not corrected), 2 (corrected).

eReferences

1. Radua J, Mataix-Cols D, Phillips ML, et al. A new meta-analytic method for neuroimaging studies that combines reported peak coordinates and statistical parametric maps. *Eur Psychiatry.* Nov 2012;27(8):605-611.
2. Viechtbauer W. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Journal of Educational and Behavioral Statistics.* 2005;30(3):261-293.
3. Viechtbauer W. Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software.* 2010;36(3):1-48.
4. Rubin DB. *Multiple Imputation for Nonresponse in Surveys.*: New York:John Wiley and Sons.; 1987.
5. Li KH, Meng XL, Raghunathan TE, Rubin DB. Significance levels from repeated p-values with multiply-imputed data. *Statistica Sinica.* 1991;1(1):65-92.
6. Rubia K, Alegria AA, Cubillo AI, Smith AB, Brammer MJ, Radua J. Effects of stimulants on brain function in attention-deficit/hyperactivity disorder: a systematic review and meta-analysis. *Biol Psychiatry.* Oct 15 2014;76(8):616-628.
7. Alegria AA, Radua J, Rubia K. (in preparation).
8. de Leeuw M, Kahn RS, Vink M. Fronto-striatal dysfunction during reward processing in unaffected siblings of schizophrenia patients. *Schizophr Bull.* Jan 2015;41(1):94-103.
9. Dowd EC, Barch DM. Pavlovian reward prediction and receipt in schizophrenia: relationship to anhedonia. *PLoS One.* 2012;7(5):e35622.
10. Nielsen M, Rostrup E, Wulff S, et al. Alterations of the brain reward system in antipsychotic naïve schizophrenia patients. *Biol Psychiatry.* May 2012;71(10):898-905.
11. Simon JJ, Biller A, Walther S, et al. Neural correlates of reward processing in schizophrenia--relationship to apathy and depression. *Schizophr Res.* May 2010;118(1-3):154-161.
12. Wotruba D, Heekeren K, Michels L, et al. Symptom dimensions are associated with reward processing in unmedicated persons at risk for psychosis. *Front Behav Neurosci.* 2014;8:382.
13. Schlagenhauf F, Sterzer P, Schmack K, et al. Reward feedback alterations in unmedicated schizophrenia patients: relevance for delusions. *Biol Psychiatry.* 2009;65:1032-1039.
14. Gradin VB, Kumar P, Waiter G, et al. Expected value and prediction error abnormalities in depression and schizophrenia. *Brain.* Jun 2011;134(Pt 6):1751-1764.
15. Morris RW, Vercammen A, Lenroot R, et al. Disambiguating ventral striatum fMRI-related BOLD signal during reward prediction in schizophrenia. *Mol Psychiatry.* Mar 2012;17(3):235, 280-239.
16. Waltz JA, Schweitzer JB, Gold JM, et al. Patients with schizophrenia have a reduced neural response to both unpredictable and predictable primary reinforcers. *Neuropsychopharmacology.* May 2009;34(6):1567-1577.
17. Wolf DH, Satterthwaite TD, Kantrowitz JJ, et al. Amotivation in schizophrenia: integrated assessment with behavioral, clinical, and imaging measures. *Schizophr Bull.* Nov 2014;40(6):1328-1337.
18. Abler B, Greenhouse I, Ongur D, Walter H, Heckers S. Abnormal reward system activation in mania. *Neuropsychopharmacology.* Aug 2008;33(9):2217-2227.
19. Diaconescu AO, Jensen J, Wang H, et al. Aberrant Effective Connectivity in Schizophrenia Patients during Appetitive Conditioning. *Front Hum Neurosci.* 2011;4:239.
20. Esslinger C, Englisch S, Inta D, et al. Ventral striatal activation during attribution of stimulus saliency and reward anticipation is correlated in unmedicated first episode schizophrenia patients. *Schizophr Res.* Sep 2012;140(1-3):114-121.

21. Gilleen J, Shergill SS, Kapur S. Impaired subjective well-being in schizophrenia is associated with reduced anterior cingulate activity during reward processing. *Psychol Med.* Jul 2014:1-12.

22. Grimm O, Heinz A, Walter H, et al. Striatal response to reward anticipation: evidence for a systems-level intermediate phenotype for schizophrenia. *JAMA Psychiatry.* May 2014;71(5):531-539.

23. Juckel G, Schlagenhauf F, Koslowski M, et al. Dysfunction of ventral striatal reward prediction in schizophrenic patients treated with typical, not atypical, neuroleptics. *Psychopharmacology (Berl).* Aug 2006;187(2):222-228.

24. Juckel G, Friedel E, Koslowski M, et al. Ventral striatal activation during reward processing in subjects with ultra-high risk for schizophrenia. *Neuropsychobiology.* 2012;66(1):50-56.

25. Koch K, Schachtzabel C, Wagner G, et al. Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. *Neuroimage.* Mar 2010;50(1):223-232.

26. Mucci A, Dima D, Soricelli A, et al. Is avolition in schizophrenia associated with a deficit of dorsal caudate activity? A functional magnetic resonance imaging study during reward anticipation and feedback. *Psychol Med.* Jan 2015:1-14.

27. Murray GK, Corlett PR, Clark L, et al. Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. *Mol Psychiatry.* Mar 2008;13(3):239, 267-276.

28. Nielsen MO, Rostrup E, Wulff S, et al. Improvement of brain reward abnormalities by antipsychotic monotherapy in schizophrenia. *Arch Gen Psychiatry.* Dec 2012;69(12):1195-1204.

29. Roiser JP, Howes OD, Chaddock CA, Joyce EM, McGuire P. Neural and behavioral correlates of aberrant salience in individuals at risk for psychosis. *Schizophr Bull.* Nov 2013;39(6):1328-1336.

30. Schlagenhauf F, Juckel G, Koslowski M, et al. Reward system activation in schizophrenic patients switched from typical neuroleptics to olanzapine. *Psychopharmacology (Berl).* Mar 2008;196(4):673-684.

31. Schlagenhauf F, Huys QJ, Deserno L, et al. Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *Neuroimage.* Apr 2014;89:171-180.

32. da Silva Alves F, Bakker G, Schmitz N, et al. Dopaminergic modulation of the reward system in schizophrenia: a placebo-controlled dopamine depletion fMRI study. *Eur Neuropsychopharmacol.* Nov 2013;23(11):1577-1586.

33. Smieskova R, Roiser J, Chaddock CA, et al. Modulation of motivational salience processing during the early stages of Psychosis. *Schizophrenia Research.* 2015:in press.

34. Walter H, Kammerer H, Frasch K, Spitzer M, Abler B. Altered reward functions in patients on atypical antipsychotic medication in line with the revised dopamine hypothesis of schizophrenia. *Psychopharmacology (Berl).* Sep 2009;206(1):121-132.

35. Waltz JA, Schweitzer JB, Ross TJ, et al. Abnormal responses to monetary outcomes in cortex, but not in the basal ganglia, in schizophrenia. *Neuropsychopharmacology.* Nov 2010;35(12):2427-2439.

36. Hägele C, Schlagenhauf F, Rapp M, et al. Dimensional psychiatry: reward dysfunction and depressive mood across psychiatric disorders. *Psychopharmacology (Berl).* Jan 2015;232(2):331-341.

37. Romaniuk L, Honey GD, King JR, et al. Midbrain activation during Pavlovian conditioning and delusional symptoms in schizophrenia. *Arch Gen Psychiatry.* Dec 2010;67(12):1246-1254.

38. Jensen J, Willeit M, Zipursky RB, et al. The formation of abnormal associations in schizophrenia: neural and behavioral evidence. *Neuropsychopharmacology.* Feb 2008;33(3):473-479.

39. Dowd EC, Barch DM. Anhedonia and emotional experience in schizophrenia: neural and behavioral indicators. *Biol Psychiatry.* May 2010;67(10):902-911.
40. Grimm O, Vollstädt-Klein S, Krebs L, Zink M, Smolka MN. Reduced striatal activation during reward anticipation due to appetite-provoking cues in chronic schizophrenia: a fMRI study. *Schizophr Res.* Feb 2012;134(2-3):151-157.
41. Harvey PO, Armony J, Malla A, Lepage M. Functional neural substrates of self-reported physical anhedonia in non-clinical individuals and in patients with schizophrenia. *J Psychiatr Res.* Aug 2010;44(11):707-716.
42. Modinos G, Tseng HH, Falkenberg I, Samson C, McGuire P, Allen P. Neural correlates of aberrant emotional salience predict psychotic symptoms and global functioning in high-risk and first-episode psychosis. *Soc Cogn Affect Neurosci.* Mar 2015.
43. Liddle PF, Laurens KR, Kiehl KA, Ngan ET. Abnormal function of the brain system supporting motivated attention in medicated patients with schizophrenia: an fMRI study. *Psychol Med.* Aug 2006;36(8):1097-1108.
44. Wolf DH, Turetsky BI, Loughead J, et al. Auditory Oddball fMRI in Schizophrenia: Association of Negative Symptoms with Regional Hypoactivation to Novel Distractors. *Brain Imaging Behav.* Jun 2008;2(2):132-145.
45. Rausch F, Mier D, Eifler S, et al. Reduced activation in ventral striatum and ventral tegmental area during probabilistic decision-making in schizophrenia. *Schizophr Res.* Jul 2014;156(2-3):143-149.
46. Rausch F, Mier D, Eifler S, et al. Reduced activation in the ventral striatum during probabilistic decision-making in patients in an at-risk mental state. *J Psychiatry Neurosci.* Jan 2015;40(1):140191.
47. Waltz JA, Kasanova Z, Ross TJ, et al. The roles of reward, default, and executive control networks in set-shifting impairments in schizophrenia. *PLoS One.* 2013;8(2):e57257.
48. Johnstone EC, Abukmeil SS, Byrne M, et al. Edinburgh high risk study--findings after four years: demographic, attainment and psychopathological issues. *Schizophr Res.* Nov 2000;46(1):1-15.