
Alan Anticevic¹,*,†, Jared X. Van Snellenberg²,‡, Rachel E. Cohen¹, Grega Repovs¹, Erin C. Dowd¹, and Deanna M. Barch¹

¹Department of Psychology, Washington University in St. Louis, St. Louis, MO 63130; ²Department of Psychology, Columbia University

These authors contributed equally to this work.

*To whom correspondence should be addressed; tel: 314-935-8459, fax: 314-935-8790, e-mail: aanticev@wustl.edu

Emotional dysfunction has long been established as a critical clinical feature of schizophrenia. In the past decade, there has been extensive work examining the potential contribution of abnormal amygdala activation to this dysfunction in patients with schizophrenia. A number of studies have demonstrated under-recruitment of the amygdala in response to emotional stimuli, while others have shown intact recruitment of this region. To date, there have been few attempts to synthesize this literature using quantitative criteria or to use a formal meta-analytic approach to examine which variables may moderate the magnitude of between-group differences in amygdala activation in response to aversive emotional stimuli. We conducted a meta-analysis of amygdala activation in patients with schizophrenia, using a bootstrapping approach to investigate: (a) evidence for amygdala under-recruitment in schizophrenia and (b) variables that may moderate the magnitude of between-group differences in amygdala activation. We demonstrate that patients with schizophrenia show statistically significant, but modest, under-recruitment of bilateral amygdala (mean effect size = −0.20 SD). However, present findings indicate that this under-recruitment is dependent on the use of a neutral vs emotion interaction contrast and is not apparent if amygdala activation by patients and controls is evaluated in a negative emotional condition only.

Key words: schizophrenia/emotion/amygdala/fMRI/meta-analysis

Since the seminal work of Bleuler¹ and Kraeplin,² emotional deficits have been considered a central component of schizophrenia symptomatology. Several authors have argued that emotional abnormalities are critical to clinical trajectory and functional outcome in this illness.³–⁶ However, “emotional processing” is not a unitary construct,⁷–⁹ and some aspects of affect processing in schizophrenia may be intact while others are abnormal.¹⁰ A recent review concluded that “in-the-moment” experience of emotion might be spared in schizophrenia¹¹; but this and other reviews⁹ have identified domains of well-documented emotional abnormalities, including (1) expression of emotion,¹²–²⁰ (2) recognition of emotional facial expressions and emotional classification,²¹–²⁵ and (3) anticipating hedonic experience.²⁶ Given behavioral evidence for some emotional deficits, there are increasing efforts to understand the neurobiology of affective disturbances in schizophrenia.¹¹,²⁷ The importance of the amygdala in affective processing of aversive stimuli is well established in healthy adults,²⁸–³² prompting many functional neuroimaging studies to focus on amygdala activation in schizophrenia.³³–³⁸ This emphasis does not rule out the involvement of other cortical or subcortical regions,³⁹ but fully understanding amygdala abnormalities may be an important starting point—particularly for processing aversive stimuli, which most consistently engage this structure.³² To date, work in this area has produced somewhat mixed findings. Since the original study by Schneider and colleagues,³³ numerous investigations have reported amygdala “under-recruitment” in patients with schizophrenia, particularly in response to aversive emotional material³⁴,³⁶,⁴⁰–⁴⁹; however, numerous other studies reported intact or even over-recruitment of the amygdala.³⁵,³⁷,³⁸,⁴⁵,⁵⁰–⁶⁶

Given these diverse findings, it is imperative to establish a quantitative summary of amygdala activation in schizophrenia—similar to work summarizing cortical involvement in other cardinal domains of dysfunction in this illness, such as dorsolateral prefrontal activation in working memory,⁶⁷,⁶⁸ episodic memory,⁶⁹ and executive function.⁷⁰ At present, there has been one attempt to conduct such a meta-analysis using activation likelihood estimation (ALE)⁷¹ to summarize studies of facial affect processing. However, studies using ALE are unable to obtain an effect size estimate of group
differences in amygdala recruitment because ALE analyzes the reliability of activation peaks across the whole brain, not the magnitude of group differences in a specific region. Additionally, ALE treats each reported peak as independent regardless of the source study; consequently, studies reporting more activation peaks have a proportionally greater impact on the results than studies reporting fewer activations, potentially yielding significant results that are driven by a small subset of the sample, or even a single study.74,75

Furthermore, ALE does not readily allow for the investigation of variables that may moderate group differences in activation, which is essential given the considerable methodological heterogeneity in this literature. For instance, 2 different task contrasts are routinely used when comparing amygdala activity between patients and controls, either a direct group comparison in the emotional condition or a group comparison for emotional minus neutral condition—ie, an interaction contrast. This difference may be critical because increased amygdala activation by patients relative to controls has been observed in response to putatively affectively neutral information.52 This implies that studies using neutral stimuli as a baseline may misinterpret increased baseline activation as under-recruitment in the emotional condition of interest. Another domain that may contribute to the heterogeneity of amygdala findings is that of individual differences in symptom severity. For example, medication types and dosages, negative and positive symptom severity, and the type of stimuli used to elicit affective processing may all impact the magnitude of this deficit, if any, across stimulus types and tasks. Moreover, given that methodological and patient variables may impact between-group differences in amygdala activation, it is important to establish whether there are moderating variables that need to be considered when interpreting results and designing studies in this field. To this end, we undertook a meta-analysis with 2 broad goals: (1) to investigate whether the literature as a whole supports the hypothesis of amygdala under-recruitment in schizophrenia in response to aversive emotional material and (2) to investigate whether there are significant moderating variables of amygdala recruitment in this illness.

Materials and Methods

Study Selection Criteria

We first identified functional neuroimaging studies utilizing an emotional task and a between-group statistical comparison of amygdala activation between patients with schizophrenia and matched controls. The Medline and PsycINFO databases were searched for articles published between Jan 01, 1998 and Aug 30, 2009, producing 855 unique results (search terms are listed in Appendix A). We included studies that (1) used either functional magnetic resonance imaging (fMRI) or positron emission tomography (PET), (2) contained both patient and control groups, (3) used a task that the authors reported contained a negative emotional manipulation, (4) conducted between-group statistical comparisons, and (5) conducted a test which detected or could have detected amygdala signal (ie either an amygdala region of interest [ROI] analysis or a whole-brain search). There were 41 studies that met these criteria.33–38,40–65,76–78

Many studies did not report a significant group difference (ie reported a null effect). One concern with such studies is that the task used may not have elicited amygdala activation even in the control group. If no amygdala signal was detected with the task, then it is difficult to argue whether a null finding reflects a true lack of a between-group difference, insufficient task sensitivity, or an underpowered study. Thus, such null effect studies were further vetted based on 2 criteria: (1) the study had to explicitly report or show that the task engaged the amygdala in either group separately or (2) the task used in the study has previously been shown to engage the amygdala based on meta-analytic work in healthy adults.29,32,79–81 Effects from 3 studies were excluded based on these criteria.49,66,82

We also identified some studies that may have activated the amygdala and detected significant group differences that, however, fell outside the scope of the present investigation because they did not employ an explicit affective manipulation83,84 or focused on a predominantly cognitive process with a minor emotional manipulation.85 This left 35 studies that were included in all analyses.

Lastly, as noted in the introduction, all the identified studies contained an aversive affective manipulation and some (15 studies) contained a separate positive manipulation. Where possible, we included only the aversive manipulations for 2 major reasons: (1) there is stronger evidence that aversive material reliably engages the amygdala in healthy adults32 and (2) the total number of studies reporting on positive manipulations was substantially smaller and thus underpowered. However, establishing a quantitative summary of between-group differences in response to positively valenced material is critical and warrants prospective investigation as more studies become available.

Selecting Moderating Variables

Next, we examined a broad set of putative moderator variables; however, the final set was constrained by the number of studies reporting information about each moderator. Thus, to balance concerns about statistical
Table 1. The Final Selection of Moderator Variables Along With Descriptive Statistics

<table>
<thead>
<tr>
<th>Variables Included in Moderator Analysis</th>
<th>Descriptives</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Family of Variables</strong></td>
<td><strong>Variable Name</strong></td>
</tr>
<tr>
<td><strong>A priori variables of interest</strong></td>
<td>Cognitive engagement on task (1) vs passive viewing (0)</td>
</tr>
<tr>
<td></td>
<td>Negative symptom severity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Medication dose in chlorpromazine equivalents</td>
</tr>
<tr>
<td><strong>Task-related variables</strong></td>
<td>Contrast of interest—group comparison for emotional condition (1) or neutral vs emotional (0)</td>
</tr>
<tr>
<td></td>
<td>Stimulus type (faces = 1 vs all other = 0)</td>
</tr>
<tr>
<td></td>
<td>Task design (blocked = 1 vs event-related = 0)</td>
</tr>
<tr>
<td></td>
<td>In-scanner behavioral performance difference (controls vs patients, effect size)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Patient characteristics</strong></td>
<td>Percent medicated (all medicated = 1 vs all unmedicated = 0)</td>
</tr>
<tr>
<td></td>
<td>Patient status (all outpatient = 1, all inpatients = 0)</td>
</tr>
<tr>
<td></td>
<td>Positive symptom severity&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>PANSS general psychopathology scale&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>BPRS or PANSS overall psychopathology&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Age at illness onset</td>
</tr>
<tr>
<td></td>
<td>Length of illness</td>
</tr>
<tr>
<td><strong>Demographic variables</strong></td>
<td>Patient gender proportion (% male)</td>
</tr>
<tr>
<td></td>
<td>Control gender proportion (% male)</td>
</tr>
<tr>
<td></td>
<td>Difference in gender proportion (controls–patients)</td>
</tr>
<tr>
<td></td>
<td>Difference in age (control vs patients, effect size)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Difference in education level (control vs patients, effect size)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Difference in IQ reported by NART or WAIS (control vs patients, effect size)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td>Difference in SES (control vs patients, effect size)&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td><strong>Imaging variables</strong></td>
<td>Amygdala ROI used (yes = 1 vs no = 0)</td>
</tr>
<tr>
<td></td>
<td>Excessive head motion verification (yes = 1 vs no = 0)</td>
</tr>
<tr>
<td></td>
<td>Hemodynamic response function model (canonical = 1 vs all other = 0)</td>
</tr>
<tr>
<td></td>
<td>Amount of smoothing (mm&lt;sup&gt;3&lt;/sup&gt;)</td>
</tr>
<tr>
<td></td>
<td>Voxel size (mm&lt;sup&gt;3&lt;/sup&gt;)</td>
</tr>
</tbody>
</table>

Note: In addition to a priori variables, we selected a number of additional moderators that were used in prior meta-analytic studies. We also selected other variables that may have moderating influence in the context of amygdala activation (eg task-related variables). The “Unit” column refers to meaning of the mean and SD values presented for each variable. PANSS, Positive and Negative Syndrome Scale; BPRS, Brief Psychiatric Rating Scale; NART, National Adult Reading Test; WAIS, Wechsler Adult Intelligence Scale; SES, socioeconomic status; ROI. Symptom severity is expressed such that a value of 1 would indicate maximal symptom reporting on a given scale by all studies. Values reflect a standardized mean difference between patients and controls (ie Hedge’s g); positive values indicate lower values for patients vs controls (eg difference in IQ level indicates that, on average, patients showed 0.77 SD lower estimated IQ when compared with controls across all studies).
power with sufficient inclusiveness, we included any variable reported by at least 12 studies. The final list of moderating variables, with summary data, is reported in table 1, and a summary of task designs and stimuli used across all included studies is presented in table 2. To facilitate correction for multiple comparisons (see below), potential moderators were grouped into 4 categories, based on differences between studies in (1) task-related variables, (2) patient characteristics, (3) imaging parameters, and (4) sample demographics. Finally, due to specific a priori hypotheses about 3 variables (severity of negative symptoms in the patient sample, the amount of antipsychotic medication being received by the patient sample, and whether participants were engaged in a cognitive task during scanning), we created a fifth family of moderators for these variables in order to enhance power for those variables for which we had an a priori hypothesis. This approach was adopted to protect against type I error for each set of conceptually related moderators, without drastically reducing power for all analyses by treating all moderators as a single family. This approach did not, in any way, reduce originally obtained sample size (as shown in table 1 for each moderator).

Table 2. List of Studies and Type of Task, Stimuli, and Emotional Rating Procedure Used

<table>
<thead>
<tr>
<th>Study</th>
<th>Publication Year</th>
<th>Task Used in Scanner</th>
<th>Stimuli Used</th>
<th>Rating Procedure (Stimuli vs Feelings)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gur et al.47</td>
<td>2007</td>
<td>Affect classification</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Rasetti et al.78</td>
<td>2009</td>
<td>Affect matching</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Hempel et al.50</td>
<td>2003</td>
<td>Affect matching and labeling</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Fakra et al.49</td>
<td>2008</td>
<td>Affect matching and labeling</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Blasi et al.54</td>
<td>2009</td>
<td>Affect matching and labeling</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Fernandez-Egea et al.65</td>
<td>2009</td>
<td>Emotion rating and gender discrimination task</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Kosaka et al.37</td>
<td>2002</td>
<td>Emotion intensity judgment</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Taylor et al.38</td>
<td>2002</td>
<td>Emotion rating</td>
<td>IAPS</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Reske et al.62</td>
<td>2009</td>
<td>Emotion labeling-happy vs sad</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Michalopoulou et al.77</td>
<td>2008</td>
<td>Gender discrimination</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Hall et al.76</td>
<td>2008</td>
<td>Gender discrimination</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Seiferth et al.63</td>
<td>2009</td>
<td>Gender discrimination</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Kang et al.59</td>
<td>2009</td>
<td>Gender discrimination</td>
<td>Auditory</td>
<td>N/A</td>
</tr>
<tr>
<td>Phillips et al.54</td>
<td>1999</td>
<td>Gender discrimination</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Surguladze53</td>
<td>2006</td>
<td>Gender discrimination</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Russel et al.55</td>
<td>2007</td>
<td>Gender discrimination (affect classification outside of scanner)</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Williams et al.48</td>
<td>2007</td>
<td>Gender discrimination (affect classification outside of scanner)</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Williams et al.43</td>
<td>2004</td>
<td>Gender discrimination (affect classification outside of scanner)</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Schneider et al.33</td>
<td>1998</td>
<td>Mood induction</td>
<td>Faces</td>
<td>Feelings</td>
</tr>
<tr>
<td>Habel et al.41</td>
<td>2004</td>
<td>Mood induction</td>
<td>Faces</td>
<td>Feelings</td>
</tr>
<tr>
<td>Reske et al.54</td>
<td>2007</td>
<td>Mood induction</td>
<td>Faces</td>
<td>Feelings</td>
</tr>
<tr>
<td>Schneider et al.56</td>
<td>2007</td>
<td>Mood induction</td>
<td>Olfactory</td>
<td>Feelings</td>
</tr>
<tr>
<td>Pauly et al.60</td>
<td>2008</td>
<td>Mood induction</td>
<td>Olfactory</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Crespo-Facorro et al.35</td>
<td>2001</td>
<td>Passive smelling</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Radulescu and Mujica-Parodi61</td>
<td>2008</td>
<td>Passive viewing</td>
<td>IAPS</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Paradiso et al.40</td>
<td>2003</td>
<td>Passive viewing</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Holt et al.51</td>
<td>2005</td>
<td>Passive viewing</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Holt et al.52</td>
<td>2006</td>
<td>Passive viewing</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Das et al.46</td>
<td>2007</td>
<td>Passive viewing-conscience unconscious (affect classification outside of scanner)</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Takahashi et al.32</td>
<td>2004</td>
<td>Passive viewing with indication of feeling</td>
<td>IAPS</td>
<td>Feelings</td>
</tr>
<tr>
<td>Dichter et al.58</td>
<td>2009</td>
<td>Target detection with aversive and neutral distraction</td>
<td>IAPS</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Johnston et al.44</td>
<td>2005</td>
<td>Tracking gender or emotion</td>
<td>Faces</td>
<td>N/A</td>
</tr>
<tr>
<td>Taylor et al.57</td>
<td>2007</td>
<td>Valence decision (valence/arousal rating outside of scanner)</td>
<td>IAPS</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Gur et al.36</td>
<td>2002</td>
<td>Valence discrimination</td>
<td>Faces</td>
<td>Stimuli</td>
</tr>
<tr>
<td>Taylor et al.45</td>
<td>2005</td>
<td>Viewing with valence judgment</td>
<td>IAPS</td>
<td>Stimuli</td>
</tr>
</tbody>
</table>

Note: IAPS, International Affective Picture System; N/A, not applicable.
Definition of Amygdala Signal
We included an effect as belonging to the amygdala if authors explicitly stated that the finding reflected a difference in amygdala recruitment or a priori amygdala ROIs were used, with one exception (one study reported amygdala activation along with Talairach coordinates that were clearly outside of the amygdala). Otherwise, we used reported Talairach coordinates and verified that the coordinates corresponded anatomically to the amygdala by using the Talairach Daemon software (http://www.talairach.org/). If the study reported Montreal Neurological Institute coordinates, we employed a transformation outlined by Brett (for more details, see Brett et al).

Calculating Symptom Severity Across Studies
To equate symptom severity across studies, we followed the procedure employed by Van Snellenberg and colleagues, which is described in Appendix B. Briefly, we converted scale scores from commonly used symptom rating scales for each study to a scale from 0 to 1 (where 0 indicated that all participants received the minimum rating on the scale and 1 indicated that all participants received a maximum rating), in order to allow for between-study comparisons.

Calculating Chlorpromazine Equivalents
For each study, we extracted medication dosages reported in chlorpromazine equivalents or converted the reported mean medication levels to chlorpromazine equivalents for studies that did not report them directly (see online supplementary material).

Effect Size Estimation
We estimated effect sizes for both amygdala activation and certain moderator variables (table 1) using an approach described by Van Snellenberg and colleagues. Briefly, we computed a standardized effect size for each study using Hedges g or estimated Hedges g from t or F statistics, P values, or published figures. All estimation procedures and formulas are presented in Appendix A.

Effect sizes were estimated from each study in the following order of preference: (1) reported means and SDs for each group or an explicit report of an effect size estimate, (2) the results of a t or an F test with one numerator df (we used test statistics or P values interchangeably), (3) Z scores reflecting a between-group activation contrast; (4) means and SEs or raw data estimated from published figures, and (5) an indication in the article of whether a significant between-group difference was observed. Specifically, if the study reported an effect as significant (but failed to provide the appropriate statistic or graph), we estimated the smallest effect required to achieve significance given the reported sample size and significance threshold. If the study conducted a test that could have detected a difference, but did not report a significant difference in amygdala activation, we estimated the effect as zero. This approach is unbiased when the true population-level effect size is 0—if the population-level effect is positive, then this approach is negatively biased and if the population-level effect is negative, then this approach is positively biased. In any case, whenever the true effect size is nonzero, this approach is conservative with respect to type I error while still allowing for the inclusion of all available data. We opted to include these studies with a zero effect size rather than exclude them because excluding them would systematically omit studies that have small effect sizes, thereby biasing the results of the meta-analysis toward larger average effects.

Effect size estimates were obtained across all studies for the left and right amygdala separately whenever possible, as well as bilaterally. Lateralized effects were estimated only from studies that explicitly reported hemisphere-specific findings, which were later averaged to obtain bilateral amygdala effect size for those studies. We used the identical approach to estimate effect sizes for other moderating variables that could be expressed as a between-group difference (eg differences in age or estimated IQ between groups).

Finally, some studies reported t or Z statistics based on the peak voxel in a cluster, while others reported the mean statistic within an ROI or cluster. Because the former approach will produce effect sizes with systematically larger absolute values than the latter, we applied a correction for studies reporting peak effects. We calculated the (unweighted) average absolute value of the effect size for each type of study, computed the ratio of mean to peak studies, and then multiplied the effect size of all peak studies by this ratio so that they would have the same expected value as studies reporting mean statistics rather than peak. A total of 10 studies reported peak values and another 10 reported means (the other studies were null effect studies). The magnitude of effect sizes based on studies reporting mean values was 0.82 times that of studies reporting peak values. Thus, we adjusted the peak values downward by this amount.

Meta-analytic Procedure
Effect sizes for each study were first adjusted for small sample bias, and the weighted mean across studies was calculated using a random-effects procedure. Rather than employing a parametric approach to calculating confidence intervals (CIs) and P values on the weighted mean, we used the bias-corrected and accelerated bootstrap (BCa bootstrap). We chose this approach because parametric procedures require the assumption of normal, independent, and identically distributed error variance and are an asymptotic solution (ie the formulas are exact
when $N$ is “large,” but how large is large enough is unknown). Specifically, because the only effect size estimate we could obtain from a number of studies was a null result, the error distribution is known to be nonnormal and nonindependent because of the large number of studies with an estimated effect size of exactly 0. In contrast, the BC$_a$ bootstrap makes no distributional assumptions and instead estimates the empirical distribution of the statistic (in this case, the weighted mean) directly, providing a CI and $P$ value based on this distribution. (While Efron and Tibshirani do not provide a means of calculating accelerated and bias-corrected $P$ values directly, this can be done by determining the CI that would have its upper or lower bound at exactly the null-hypothesis value being tested. Thus, one simply observes the proportion of the bootstrap distribution falling below the null-hypothesis value of the statistic and applies the function inverse of the BC$_a$ correction given for CIs to this proportion.) Similarly, putative moderator variables were analyzed in the weighted-least squares regression procedure of Hedges and Olkin, but CIs and $P$ values on the parameters were obtained from the BC$_a$ bootstrap.

Model selection for moderator analysis was carried out in a step-forward fashion and was done separately for left, right, and bilateral amygdala activation differences. That is, at the first step, each of the possible single-parameter models was estimated and the model with the lowest $P$ value was selected. At the second step, each of the possible 2-parameter models was estimated (including the moderator selected at step one), and again the model with the lowest $P$ value was selected. At each step, a new moderator was included if its $P$ value was less than .10. If the $P$ value of any moderator included at an earlier step became larger than .15 it was removed from the model. Once the final model was selected, $P$ values and CIs were obtained for all putative moderators as compared with this model. A flowchart of the entire analysis approach is presented in figure 1.

The multiple comparison problem in moderator analyses was dealt with by using false-discovery rate (FDR) correction at $P < .05$ within each of 5 families of moderators (see table 1).

All analyses were carried out on a single set of 100 000 bootstrap resamplings of the original data. Two included studies reported data on more than one contrast that met our inclusion criteria and thus had more than one estimated effect size (ie multiple task conditions). Whenever one of these studies was selected in a bootstrap sample, one of the possible estimated effect sizes was randomly selected. This approach was taken rather than taking a mean or median of the available effects for each study prior to bootstrapping because it more appropriately models the actual variability in the observed data. However, this approach makes ambiguous what constitutes the “real” observed data set, which is required for the BC$_a$ procedure. That is, because more than one effect size is reported in some studies and these effects cannot be treated as if they came from separate studies because of nonindependence, it is unclear which of the effect sizes is the “true” observed value for that study. Indeed, there are actually 4 possible “real” data sets based on all the reported effects (because 2 studies reported 2 usable effect sizes). Consequently, we carried out the BC$_a$ procedure for each of these 4 data sets, and all reported effect size estimates, parameter estimates (for moderators), CIs, and $P$ values are the median value from these 4 possibilities. All reported $P$ values are 2-tailed.

Fig. 1. Schematic Representation of the Meta-analytic and Moderator Model Selection Procedures.
Results

Effect Size Estimates

A forest plot of the effect size estimates for all included studies is presented in figure 2. Across studies, patients with schizophrenia exhibited significantly reduced activation of bilateral amygdala (mean effect size = −0.22; 95% CI = −0.37 to −0.08; \( P = .002 \)) and right amygdala (mean = −0.17; 95% CI = −0.37 to −0.03; \( P = .012 \)). An effect in the same direction was observed for left amygdala but did not reach conventional levels of statistical signifi-
cance (mean = −0.13; 95% CI = −0.31 to 0.04; \( P = .136 \)). The bootstrapped data and 95% confidence intervals for these 3 analyses are shown in figures 3A–C.

Moderator Analyses

The model selection procedure for group differences in bilateral amygdala activation resulted in a model includ-
ing 3 potential moderators: “task contrast of interest” (a categorical variable of whether the effect size for a given study was derived from a direct group comparison in the emotional condition or from a group comparison using the contrast of neutral vs emotional conditions), “voxel size” (mm\(^3\)), and “excessive head motion verification.” However, only task contrast of interest was significant after FDR correction (\( P = .009 \)); the parameter estimate was 0.35 (95% CI = 0.09–0.70), indicating an estimated 0.35 SD increase in amygdala activation by patients relative to control participants when groups are compared directly in the emotion condition rather than in a neutral vs emotion contrast. Excessive head motion verification reached traditional levels of statistical significance prior to FDR correction (\( P = .016 \)); the parameter estimate was −0.29 (95% CI = −0.53 to −0.06), indicating an estimated 0.29 SD decrease in amygdala activation by patients relative to control participants in studies that
employed a check for excessive head motion. Finally, voxel size did not achieve traditional levels of statistical significance ($P = .127$).

As a result of the finding that the task contrast used across studies had a significant impact on differences in amygdala activation across groups, we carried out an a posteriori follow-up analysis to determine whether there was any evidence of reduced activation of amygdala in studies using only a between-group comparison in the emotion condition rather than a neutral vs emotion contrast. We simply repeated the effect size estimation procedure for bilateral amygdala reported above but calculated separately for studies using a direct comparison and those using a contrast. Indeed, for studies using a direct comparison, there was no evidence of amygdala under-activation by patients (mean effect size = −0.04; 95% CI = −0.17 to 0.12; $P = .688$), while for studies using a neutral vs emotion contrast, there was considerable

---

**Fig. 3.** Bootstrap Distributions of the Weighted Mean Effect Size (Hedge’s $g$) for Control vs Patient Difference in Amygdala Activation Shown for: (A) Bilateral, (B) Right amygdala, and (C) Left amygdala. Dotted lines indicate 95% CIs, and the thicker dashed line indicates the estimated mean effect size across studies.
Evidence for reduced amygdala activation by patients (mean effect size = −0.36; 95% CI = −0.54 to −0.23; \( P < .001 \)) (figure 4).

The model selection procedure for group differences in right amygdala activation resulted in a model including 2 potential moderators, “task contrast of interest” and “amygdala ROI used.” Only amygdala ROI used was significant after FDR correction (\( P < .001 \)); the parameter estimate was 0.34 (95% CI = 0.14–0.62), indicating an estimated 0.34 SD increase in activation of the right amygdala by patients relative to control participants in studies employing an amygdala ROI rather than a whole-brain comparison. Task contrast of interest was not significant following FDR correction but did reach traditional levels of statistical significance prior to FDR correction (\( P = .040 \)); the parameter estimate was 0.26 (95% CI = 0.01–0.52), indicating an estimated 0.26 SD increase in amygdala activation by patients relative to control participants for group comparisons made directly in the emotion condition (similar to the results for bilateral amygdala, above).

The model selection procedure for group differences in left amygdala activation resulted in a model including 2 potential moderators, “age at illness onset” and “group difference in education.” Neither moderator was significant after FDR correction, although group difference in education reached traditional levels of statistical significance prior to correction (\( P = .010 \)); the parameter estimate was 0.40 (95% CI = 0.16–1.11), indicating an estimated 0.4 SD increase in amygdala activation by patients for every SD increase in control subject’s years of education relative to patients.

Discussion

Results of the present investigation demonstrate that, across all neuroimaging studies using a negatively valenced emotional manipulation, there is evidence for slightly reduced amygdala activation (approximately one-fifth of a SD) bilaterally by patients with schizophrenia vs controls. However, moderator analysis and an a posteriori follow-up analysis demonstrate that amygdala under-recruitment is present only in studies that employed a neutral vs emotion contrast and not in studies directly comparing patients and controls in the emotion condition.

These findings suggest that the apparent deficit in amygdala activation during negative emotional states in patients may be due to elevated amygdala responses to emotionally neutral stimuli, consistent with the results of an earlier qualitative review and imaging work using neutral stimuli. Thus, patients with schizophrenia appear to have a normal amygdala response (relative to a resting baseline) to affectively aversive stimuli of the type commonly used in the neuroimaging literature but may have elevated amygdala responses to emotionally neutral stimuli—resulting in an apparent deficit in activation when a neutral vs emotion interaction contrast is employed. Importantly, in the present meta-analysis, we were unable to directly examine amygdala responsiveness in the neutral condition due to the small number of studies explicitly reporting or testing for this effect. However, as studies start to conduct comparisons of the neutral condition in isolation, prospective work should investigate this possible amygdala abnormality in schizophrenia.

Amygdala Recruitment in Schizophrenia

Across all studies, we found a small but significant reduction in amygdala activation by patients with schizophrenia for bilateral amygdala, as well as in the right amygdala. While the reduction in amygdala activation by patients observed in the left amygdala did not reach the traditional threshold for statistical significance, the 95% CI for left amygdala overlapped extensively with the CIs for bilateral and right amygdala, suggesting that under-recruitment is present in the amygdala bilaterally.

However, as noted above, the type of contrast used significantly moderated this finding. That is, patients showed negligible differences in amygdalar responsiveness when directly compared with controls for the emotionally aversive condition specifically, but showed considerable amygdala under-recruitment when studies used an emotion minus neutral contrast. An earlier qualitative review concluded that patients showed consistent reductions in amygdala activation when emotional stimuli were compared with neutral but not when groups were compared directly in the emotionally evocative condition, an observation that now has direct empirical support. Also, present findings are highly consistent with prior work suggesting intact in-the-moment experience...
of affective stimuli in schizophrenia. Thus, patients may show abnormal amygdala activation to affectively neutral events (i.e., the control condition), as observed in prior work, suggesting aberrant responses to environmentally nonsalient events. Equally important, this result highlights that design and analysis considerations (such as choice of contrast condition) are crucial because they will allow consistent replication and comparison across studies.

However, due to the course spatial and temporal resolution reported by most studies, we focused on the entire amygdalar complex. It may still be possible that specific amygdala subnuclei manifest unique patterns of pathology and under-recruitment even for a direct comparison in the emotional condition. Additionally, studies employing more exquisite temporal resolution may find signal abnormalities in amygdala time courses in response to emotional stimuli. Furthermore, it may be possible that even though the amygdala shows intact response amplitude in patients, it may still exhibit abnormalities in between-region functional interaction, which may be explored with functional connectivity analyses in the future. That is, the amygdala does not process survival-relevant information in isolation but functions as part of a broader circuit involved in the detection of emotional salience and activation seen in the amygdala must be interpreted with this in mind. Therefore, current findings are limited in interpretation given that amygdala activation was not considered along with other regions involved in emotional processing. For instance, investigating amygdala activation in the context of a broader circuit will be especially critical for specific aspects of emotional processing where patients have well-documented deficits and may involve the amygdala’s interaction with other regions (e.g., emotional facial expression identification).

Additional Moderator Variables

The only other significant moderator finding was that the magnitude of the between-group difference in right amygdala was smaller if a study employed ROI-based analyses. Registration inequality across clinical and control groups can critically impact the pattern of between-group differences. One possibility is that using anatomically or functionally delineated ROIs at an individual subject level minimizes the impact of across-subject movement and anatomical variability. That is, if a study used individual subjects’ anatomy to extract amygdala signals, then differences in registration quality no longer present a problem and may result in higher signals in the patient group. Similarly, well-defined functional masks may better capture peak amygdala signal in each group, thus increasing the probability of accurately assaying signal levels in patients. We acknowledge that we had no a priori predictions for a lateralized effect with regard to ROI use. However, it may be possible that, given some evidence for weaker right amygdala responsiveness to affectively negative materials, using ROIs may aid signal detection in this region where statistical power may be compromised relative to the left amygdala—an important consideration for future studies aimed at detecting activation in this region.

It is important to note that although our analyses of other moderator variables did not produce significant results, none of the variables we included in our analysis can be ruled out as potential moderators of between-group differences in amygdala activation during negative emotional processing. The number of studies on this topic is still relatively small; thereby limiting our power to detect small but potentially important effects, and this problem is compounded by the fact that not all studies reported data on many of the moderators of interest. Consequently, during the model selection process as new variables are included in a model power is diminished not only because of the presence of additional parameters in the model but because the number of observations is reduced due to missing data. The literature would likely benefit substantially from more complete reporting of variables that may influence amygdala activation differences between control participants and patients with schizophrenia so that future meta-analytic work can more thoroughly characterize which variables are important moderators of these differences.

Valence

We focused explicitly on studies and task contrasts employing an aversive emotional manipulation. The main reason for this approach was statistical power because a much smaller subset of studies contained a positive manipulation. As studies accumulate, future meta-analyses should attempt to synthesize amygdala findings in schizophrenia in studies of positive emotion. Such work will aid our understanding of amygdala involvement in perceiving pleasant sensory events and, in turn, help better understand the amygdala’s role in anhedonia pathology seen in this illness. Also, present findings suggest that patients may show aberrant amygdala responsiveness to affectively neutral stimuli. Thus, a logical implication is that a meta-analysis should be conducted examining this effect. At present, due to the same hurdles mentioned for positive valence effects, such a meta-analysis was not attempted; however, it will be critical to directly examine this putative abnormality as prospective investigations become available.

Limitations

While we made our best attempt to obtain a quantitative estimate of amygdala under-recruitment in
schizophrenia, there are limitations that should be taken into account. The available data did not allow for more spatial and temporal precision with regard to examining functional alterations in specific amygdala subdivisions in schizophrenia. In addition, the analysis utilized a total of 35 studies, with occasionally fewer studies available for specific moderators. Thus, it may be possible that certain moderator variables were underpowered to detect smaller effects due to restricted sample of available studies reporting on that moderator. It will be critical, as the literature continues to grow, to revisit the present moderator analyses and conduct an even finer-grained investigation of moderation effects focusing on additional moderators not addressed at present (eg reporting signal-to-noise ratio for each group in the amygdala).101,102

Conclusion
We demonstrated that, across studies, patients with schizophrenia show significant amygdala under-recruitment in response to emotionally aversive material, but that this finding is qualified by the nature of the task contrast: No significant difference emerged when a direct group comparison was carried out for the emotion condition. We also showed that task design and use of a priori ROIs are vital when examining amygdala activation abnormalities in schizophrenia, allowing consistent replication across studies. Overall, present findings suggest that amygdala recruitment in schizophrenia is intact when groups are directly compared in an emotionally aversive condition but that a deficit is apparent when an emotional vs neutral contrast is employed, suggesting possible abnormalities in perceiving affectively neutral information in this illness.

Supplementary Material
Supplementary material is available at http://schizophreniabulletin.oxfordjournals.org.

Funding
Dr Barch has received grants from the National Institute of Mental Health, National Institute on Aging, National Alliance for Research on Schizophrenia and Depression, Novartis, and the McDonnell Center for Systems Neuroscience.

Acknowledgments
We thank T. Yarkoni and 2 anonymous reviewers for their helpful comments and suggestions.

Appendix A

Database search terms: schizophren* AND (fMRI OR neuroimaging OR functional neuroimaging OR PET) AND (emotion OR affect OR affective OR emotional)

\[
\text{Hedge's } g = \frac{\bar{x}_c - \bar{x}_t}{\sqrt{\left(\frac{(n_c - 1)x_c + (n_t - 1)x_t}{n_c + n_t - 2}\right)}} \quad (1)
\]

\[
\text{Hedge's } g = t\sqrt{\left(\frac{n_t + n_c}{n_t n_c}\right)} \quad (2)
\]

\[
\text{Hedge's } g = F\left(\frac{n_t + n_c}{n_t n_c}\right) \quad (3)
\]

Where subscripts \(c\) and \(t\) refers to control and patient samples, respectively; \(x\) sample mean; \(s\), \(SD\); \(n\), sample size; \(t\), \(F\) inferential sample statistics (\(F\) refers only to \(F\) statistics with 1 numerator \(df\)). Statistical parametric mapping \(Z\) scores of the maximum activation difference between patients and controls were converted to \(P\) values. Estimates from \(P\) values were made by obtaining the \(t\) statistic corresponding to the reported \(P\) on a \(t\) distribution with \((n_c + n_t - 2) df\), which was then converted to Hedge’s \(g\) based on formula (2).

Where no mean or statistic was directly reported in the main text or table, we used the published figures to extract the relevant mean statistics. Estimates from published figures were made from high-resolution screenshots of figures in PDF versions of articles. A 1-pixel grid was overlaid, and a straight line was used to measure the precise height of means and SEs in the figure, which were then converted to Hedge’s \(g\) based on formula (1).

Appendix B

As noted above, symptom severity was calculated to rescale the values and standardize across all studies on a scale from 0 (no symptoms) to 1 (maximal symptoms on a given scale). To obtain this score, we divided the mean score reported for a sample by the maximum possible score on the scale (with an adjustment for scales with a minimum possible scores of 1 vs 0).

We calculated separate scores for (1) positive symptoms using Positive and Negative Syndrome Scale (PANSS)\textsuperscript{103} positive score or Scale for the Assessment of Positive Symptoms (SAPS)\textsuperscript{104} score, (2) negative symptoms using PANSS negative score or Scale for the Assessment of Negative Symptoms (SANS)\textsuperscript{104} score, and (3) overall pathology using the Brief Psychiatric Rating Scale (BPRS)\textsuperscript{105} or PANSS total score. If none of these total scores were available, we used the average of the SAPS and SANS scores, average of PANSS
positive and PANSS negative scores, or PANSS general psychopathology scale. It should be noted that symptom severity was defined at the study level, thus present results only provide sample-to-sample variability in symptoms.

For example, if scale range for an individual item was started at 1 then:

\[
\text{Symptoms} = \frac{(\text{RS} - N_{\text{items}})}{(\text{UR} - 1) \times (N_{\text{items}})}
\]

Where \( \text{RS} \) = reported score for a given study, \( N_{\text{items}} \) = total number of questions on the scale used for that study, \( \text{UR} \) = upper range on a given item for the used scale. For instance, Brief Psychiatric Rating Scale (BPRS) measures overall psychopathology and is comprised 18 items rated on a 1–7 Likert scale. The maximal possible score on this scale is 126, and the minimal possible score is 18. Therefore, if a study reported a BPRS rating of 18, based on equation 4 that would translate into symptom severity of 0 (ie minimal possible score and absence of symptoms). In contrast, if a study reported a maximal symptom rating of 0 (ie minimal possible score and absence of symptoms). The middle point between 18 and 126 is 72, thus if a study reported mean symptom severity of 72 that would translate into 0.5 (50% of maximal symptom expression on BPRS).

References


