The Lay Assessment of Subclinical Depression in Daily Life

Matthias R. Mehl
University of Arizona

This study examined how laypersons assess subclinical depression in others on the basis of information about their daily lives. For 2 days, 96 participants were tracked with the Electronically Activated Recorder, a naturalistic observation method that samples ambient sounds from participants’ momentary environments. Judges rated participants’ levels of depression after listening to the sampled ambient sounds. Participants’ depressive symptoms were assessed with the Beck Depression Inventory. Overall, judges showed little accuracy at determining participants’ levels of depressive symptoms from the ambient sounds. Exploratory analyses, however, revealed that judges were more accurate among moderately and severely depressed participants, presumably because the cues judges used to assess depression (e.g., spending time alone, not socializing, not laughing) discriminated successfully only at high levels of subclinical depression.

Keywords: peer assessment, manifestations of depression, person perception, personality judgment, Electronically Activated Recorder

The clinical assessment of depression is of paramount importance in psychology. Yet, in daily life, it is usually not expert clinicians but peers that make initial judgments of the extent to which an individual is depressed. Such lay assessments have important clinical implications given the role that peers play in providing social support and recommending lay and professional treatment options. Empirically, though, lay assessments of depression have received comparatively little attention. This study pursued two major goals: First, it sought to determine how accurately laypersons assess subclinical depression in others based on information about their daily lives. Second, it explored the validity of potential perceptual cues that may be involved in daily life-based lay assessments of subclinical depression.

How accurate would we expect laypersons to be at detecting depression in others? Two different perspectives inform this question, but they suggest two different answers. From a clinical perspective, depression is a psychological disorder for which research has confirmed deficits in social and emotional functioning such as impaired social skills and disturbed interpersonal communications (e.g., Coyne, Burchill, & Stiles, 1990; Rottenberg & Gotlib, 2000; Segrin, 2000). The interpersonal nature of these deficits suggests that depression should be quite readily detectable from people’s social encounters. Consequently, from a clinical perspective, lay assessments of depression should be characterized by high degrees of accuracy. In line with this notion, Coyne (1976) found that laypersons recognized and differentially reacted to depressed and nondepressed targets on the basis of a brief informal phone conversation. From a personality perspective, subclinical depression is considered an emotional trait. Research has shown that trait depression and its conceptual relative, trait negative emotionality, are private traits with predominantly experiential and few behavioral correlates (e.g., Borkenau & Liebler, 1992; Nezlek, Hampton, & Shean, 2000; Nezlek, Imbrie, & Shean, 1994). Consequently, from a personality perspective, lay assessments of subclinical depression should be characterized by low degrees of accuracy. In line with this notion, several person perception studies have shown that compared with highly observable traits such as extraversion or conscientiousness for which minimal information is sufficient to bring about accurate personality judgments, lay assessments of emotional traits such as negative emotionality or neuroticism generally show little accuracy (e.g., John & Robins, 1993; Watson, Hubbard, & Wiese, 2000).

A possible explanation for the discrepant findings is that depressive symptoms in normal populations are qualitatively different from processes in clinical depression. A different explanation assumes that depression is dimensional in nature (Gotlib, Lewinsohn, & Seeley, 1995) and that clinical studies sample from more severely depressed populations than personality research. This explanation, however, implies that depression becomes detectible to observers only once it reaches a critical severity. Consistent with this notion, Nezlek et al. (1994) found that depression was related to fewer social interactions only among highly depressed persons.

Such discontinuity in behavioral manifestations, however, raises important questions for the lay assessment of subclinical depression: Are laypersons aware of this discontinuity, that is, do they use different cues when rating severely versus non- or mildly depressed individuals? Do inaccurate lay assessments result from...
the use of nondiagnostic cues or from an overgeneralized use of cues that are diagnostic only among highly depressed individuals?

Overview of the Study and Research Questions

This study examined the process and accuracy of daily life-based lay assessments of subclinical depression. Participants’ real-world social encounters were tracked for 2 days with the Electronically Activated Recorder (EAR; Mehler, Pennebaker, Crow, Dabbles, & Price, 2001), a naturalistic observation method that samples snippets of ambient sounds from people’s momentary environments. With its high-frequency sampling (roughly 70 clips per day), the EAR provides an acoustic log of a person’s day as it naturally unfolds (Mehler & Pennebaker, 2003a, 2003b). The sampled ambient sounds were then played to naïve judges who rated participants’ levels of depression on the basis of their naturally-occurring daily social behaviors and interactions.

Brunswik’s (1956) lens model was used to conceptualize the relationship between participants’ social encounters, the judges’ ratings of participants’ depression, and participants’ actual levels of subclinical depression. According to Brunswik (1956), cues in the environment serve as lenses through which observers perceive psychological constructs. For example, social withdrawal (e.g., not going out) may serve as a lens through which observers perceive targets’ depression. In Brunswik’s lens model, the probabilistic link between an observable cue (e.g., not going out) and the observers’ perceptions (e.g., level of depression) is referred to as cue utilization. Cue validity refers to the probabilistic link between the observable cue and the targets’ actual standing on the psychological construct (e.g., self-reported depressive symptoms). This degree that both links match, the observers’ perceptions should converge with the psychological construct being observed. This convergence—referred to by Brunswik (1956) as functional achievement—can be considered an indicator of judgmental accuracy (Wiggins, 1973).

The study used the lens model framework to address the following three research questions: (1) How accurately can laypersons assess subclinical depression from people’s daily lives? (2) What cues do laypersons use when they assess depression in others (cue utilization)? (3) How diagnostic are these cues of people’s actually experienced levels of depressive symptoms (cue validity)? The study finally explored the extent to which lay assessments of subclinical depression depended on the severity of the targets’ levels of depressive symptoms.

Method

For this study, 96 introductory psychology students (47 women, 49 men; mean age = 18.7 years) served as target participants in exchange for course credit. The EAR system consisted of a digital voice recorder (SONY ICD-MS1, Sony Corporation of America, New York, New York), an external tie clip microphone (OPTIMUS Tie Clip Microphone, OPTIMUS, Fort Worth, Texas), and a controller chip (JohnPrice, http://www.tiwire.com; Mehler et al., 2001). The chip was programmed to produce 4.8, 30-s recordings per hour. Participants wore the EAR for 2 weekdays during their waking hours. They were thoroughly informed about the study’s privacy and confidentiality policies and had an opportunity to erase recordings they did not want the researchers to hear. Of the participants, 19 used this opportunity, and 3 erased a total of 10 sound files. On average, participants provided 130 (SD = 34) valid waking sound files (~65 min).

Eight research assistants coded all sound files for aspects of participants’ locations, activities, interactions, and moods (see Mehler, Gosling, & Pennebaker, in press). Each participant was coded by one research assistant. This study focused on a subset of variables with strong theoretical connections to depression (e.g., talking, socializing, laughing; see Table 1). Interencoder agreement (ICC[2,k]; Shrout & Fleiss, 1979), determined from a set of training EAR recordings (392 sound files), independently coded by all eight research assistants, exceeded .85 for all categories.1 The raw binary codings were converted into time-use estimates by calculating the percentage of a person’s waking EAR recordings in which a coding category applied. Research assistants further transcribed all of the participants’ utterances captured by the EAR. The transcripts were submitted to Linguistic Inquiry and Word Count (LIWC; Pennebaker, Francis, & Booth, 2001), a word count-based text analysis program. For this study, only LIWC categories with strong theoretical connections to depression were selected (e.g., sadness, optimism words; see Table 1). Intertranscriber agreement (ICC[2,k]) exceeded .85 for all categories.

Participants’ depressive symptoms were assessed with the 13-item short form of the Beck Depression Inventory (BDI–SF; Beck & Beck, 1972) during the initial session prior to the EAR monitoring. Correlations between the BDI–SF and the full 21-item version range between .89 and .97 (Beck, Rial, & Rickels, 1974). The scale’s internal consistency was high ($\alpha = .88$); participants’ scores ranged from 0 to 23 ($M = 3.76, SD = 4.72$, possible range = 0–39). On the basis of the recommended cutoff scores, I put 13 participants in the mild range (scores between 5 and 7), 11 in the moderate range (scores between 8 and 15), and 3 in the severe range (scores $\geq 16$).

A second group of 18 research assistants (11 women, 7 men, mean age = 21.0 years) served as naïve judges of the target participants’ levels of depression. The 18 judges were divided into three groups and each target participant was rated on average by 6 judges.2 After listening to each participant’s recordings, judges rated the person on a number of characteristics, including depression, the Big Five personality dimensions, and various other traits (Mehler et al., in press). Specifically, judges answered the question “I see the person as someone who is depressed” on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). Interrater reliability (ICC(1,k)) was high ($r = +.71, p < .01$). Idiosyncrasies in scale use were removed by standardizing the ratings within judge before aggregating them across judges.

Results

The three research questions were addressed within three sets of analyses: First, the overall accuracy of judges’ lay assessments was determined by correlating the aggregated judges’ ratings with participants’ BDI–SF scores. Second, judges’ cue utilization was analyzed by correlating the aggregated judges’ ratings with the EAR-derived information about participants’ daily lives.3 Third, the validity of the daily-life cues was analyzed by correlating the EAR-derived cues with participants’ levels of depressive symptoms. Finally, to explore the extent to which judges’ lay assessments depended on the level of participants’ depressive symptoms,
Table 1

Lens Model Analysis of Judges’ Lay Assessments of Subclinical Depression: Correlations Between EAR-Derived Cues and Judges’ Ratings of Participants’ Depression (Cue Utilization) as Well as Participants’ Self-Ratings of Depressive Symptoms (Cue Validity)

<table>
<thead>
<tr>
<th>SECSI/LIWC category</th>
<th>Entire sample</th>
<th>BDI-SF &lt; 8</th>
<th>BDI-SF ≥ 8</th>
<th>Entire sample</th>
<th>BDI-SF &lt; 8</th>
<th>BDI-SF ≥ 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interactions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alone</td>
<td>63.5, 16.2</td>
<td>−0.03</td>
<td>0.04</td>
<td>0.34</td>
<td>0.60**</td>
<td>0.56**</td>
</tr>
<tr>
<td>Talking</td>
<td>34.5, 13.3</td>
<td>−0.05</td>
<td>0.00</td>
<td>−0.33</td>
<td>−0.52**</td>
<td>−0.45**</td>
</tr>
<tr>
<td>% dyadic conversations</td>
<td>58.3, 17.1</td>
<td>−0.06</td>
<td>−0.07</td>
<td>0.37</td>
<td>−0.01</td>
<td>−0.12</td>
</tr>
<tr>
<td>% group conversations</td>
<td>26.6, 18.0</td>
<td>0.17</td>
<td>0.03</td>
<td>−0.31</td>
<td>−0.32**</td>
<td>−0.30**</td>
</tr>
<tr>
<td>Activities</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Entertainment</td>
<td>4.9, 9.6</td>
<td>−0.01</td>
<td>0.10</td>
<td>−0.41</td>
<td>−0.21*</td>
<td>−0.14</td>
</tr>
<tr>
<td>Socializing</td>
<td>12.3, 13.1</td>
<td>0.11</td>
<td>−0.18</td>
<td>−0.27</td>
<td>−0.35**</td>
<td>−0.29**</td>
</tr>
<tr>
<td>Mood</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Laughing</td>
<td>8.2, 5.4</td>
<td>−0.02</td>
<td>−0.19</td>
<td>−0.18</td>
<td>−0.40**</td>
<td>−0.37**</td>
</tr>
<tr>
<td>Crying</td>
<td>0.1, 0.3</td>
<td>0.04</td>
<td>0.07</td>
<td>0.11</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td>Language use</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sampled raw word count</td>
<td>1004.0, 521.0</td>
<td>−0.12</td>
<td>−0.07</td>
<td>−0.49</td>
<td>−0.48**</td>
<td>−0.40**</td>
</tr>
<tr>
<td>1st person singular pronouns</td>
<td>6.7, 1.7</td>
<td>0.20*</td>
<td>0.08</td>
<td>0.80**</td>
<td>0.17</td>
<td>0.10</td>
</tr>
<tr>
<td>Positive emotion words</td>
<td>2.3, 0.8</td>
<td>0.00</td>
<td>−0.09</td>
<td>−0.17</td>
<td>−0.10</td>
<td>−0.13</td>
</tr>
<tr>
<td>Optimism</td>
<td>0.2, 0.2</td>
<td>−0.13</td>
<td>0.10</td>
<td>−0.36</td>
<td>−0.14</td>
<td>−0.07</td>
</tr>
<tr>
<td>Negative emotion words</td>
<td>1.6, 0.9</td>
<td>0.12</td>
<td>0.07</td>
<td>0.06</td>
<td>−0.26**</td>
<td>−0.24*</td>
</tr>
<tr>
<td>Sadness words</td>
<td>0.2, 0.2</td>
<td>−0.01</td>
<td>0.03</td>
<td>0.31</td>
<td>−0.05</td>
<td>−0.06</td>
</tr>
<tr>
<td>Anger words</td>
<td>0.8, 0.8</td>
<td>0.02</td>
<td>0.15</td>
<td>−0.27</td>
<td>−0.30**</td>
<td>−0.22*</td>
</tr>
<tr>
<td>Causation words</td>
<td>1.2, 0.6</td>
<td>0.02</td>
<td>0.01</td>
<td>0.53*</td>
<td>0.24*</td>
<td>0.22*</td>
</tr>
</tbody>
</table>

Note: N = 96. Participants’ depressive symptoms were assessed with the short form of the Beck Depression Inventory (BDI–SF). Judges rated participants’ levels of depression on the basis of a single item. BDI–SF < 8 = participants below the cutoff point for moderate depressive symptoms (n = 82); BDI–SF ≥ 8 = participants above the cutoff point for moderate depressive symptoms (n = 14). The Electronically Activated Recorder (EAR)-derived cues were computed as proportions of the total number of sampled sound files (Interactions, Activities, Mood) or words (Language Use; except sampled raw word count). % variables are computed as proportion of participants’ total amount of talking. SECSI = Social Environment Coding of Sound Inventory; LIWC = Linguistic Inquiry and Word Count analysis.

* p ≤ .05, two-tailed. ** p ≤ .01, two-tailed.

Question 1: Accuracy of Judges’ Lay Assessments

Figure 1 shows the scatter plot of judges’ lay assessments. Across the entire sample, the aggregated judges’ ratings were not significantly correlated with participants’ BDI–SF scores (r = .13, p = .23, 95% confidence interval [CI] = −.08, .32), indicating that overall, judges showed little accuracy in determining participants’ levels of depressive symptoms from the sampled ambient sounds. The subgroup analyses, however, qualified this overall null finding: Whereas among non- and mildly depressed participants, judge-rated depression was entirely unrelated to participants’ BDI–SF scores (r = .05, 95% CI = −.17, .26), a considerably higher level of accuracy emerged among moderately or severely depressed participants (r = .55, p = .04; 95% CI = .03, .84).4

In a follow-up analysis, the reliability of this subgroup difference was estimated with the original raw (i.e., nonstandardized) and nested (i.e., nonaggregated) data. A two-phase mixed-effect multilevel regression analysis with target participants as Level 1 units, judges as Level 2 units, and fixed discontinuities in intercept and slope at the critical cutoff point (BDI–SF = 8; cf. Singer & Willett, 2003, Chapter 6) yielded both a significant discontinuity in intercept, b = −1.12, t(570) = −2.53, p = .01, and slope, b = 0.08, t(570) = 1.81, p = .04, one-tailed. The discontinuity in slope indicates that the association between participants’ BDI–SF scores and the judges’ ratings of participants’ levels of depression was significantly stronger above, b = 0.10, compared with below, b = 0.02, the cutoff point.

4 With a BDI–SF score of 23 (z = 4.4), the most depressed participant was a univariate outlier in the sample (see Figure 1). This created the dilemma that, because of the small number of moderately and severely depressed participants, excluding the participant from the analyses could critically affect the study findings; on the other hand, including the participant in the analyses seemed critical to maximize their generalizability. Analyses without the participant yielded a slightly reduced correlation for judges’ overall accuracy (r = .01). They further yielded a reduced and now no longer statistically significant accuracy correlation among moderately and severely depressed participants (r = .24, p = .22, one-tailed). This reduced level of accuracy, however, was still larger than the one obtained among the non- or mildly depressed subgroup (r = .05). Finally, excluding the participant from the analyses did generally not reduce and did in some instances even slightly increase the magnitude of both the cue-utilization and the cue-validity correlations among moderately and severely depressed participants. For example, among participants with BDI–SF scores ≥ 8 (n = 13), judges’ perceptions of participants’ depression correlated .82 with the time spent alone; −.64 with the relative amount of group conversations; −.52 with the time spent on entertainment; and −.66 with the time spent socializing. Similarly, among participants with BDI–SF scores ≥ 8, participants’ depressive symptoms correlated .28 with the time spent alone; −.47 with the relative amount of group conversations; −.44 with the time spent on entertainment; and −.31 with the time spent socializing. Taken together, this suggests that the main study conclusions are independent of whether the most depressed participant is dropped from the analyses.
Question 3: Cue Validity

The left hand side of Table 1 shows how participants’ levels of depressive symptoms were actually related to observable aspects of their daily lives. In line with the lack of behavioral manifestations of depression found in other studies (e.g., Nezlek et al., 2000; 1994), across the entire sample, participants’ BDI–SF scores were unrelated to the time they spent alone, talking, on entertainment, socializing, and behaviorally or verbally expressing positive and negative emotions. Consistent with prior research on self-awareness processes in depression (Rude, Gortner, & Pennebaker, 2004), a linguistic self-focus as indicated by frequent self-referencing emerged as the only cue diagnostic of participants’ levels of depressive symptoms.

The findings for the subgroups again yielded a more differentiated picture: No significant correlations emerged for participants with no or few depressive symptoms. Among those with moderate and severe depressive symptoms, however, a pattern of behavioral manifestations emerged that closely corresponded to judges’ cue utilization: The more severe participants’ symptoms were, the more time they spent alone, the less they talked, particularly in groups, the less they socialized and engaged in entertainment, and the less they laughed. Depressive symptoms were also related to use of first person singular, optimism, sadness, and causation words (e.g., why, because; linguistic markers of meaning-making concerns). Unfortunately, because of the small sample size, many individual correlation coefficients failed to meet conventional significance thresholds. Yet, the emerging pattern of manifestations at high levels of depressive symptoms is consistent with previous research (Nezlek et al., 1994) and with judges’ more successful discrimination among degrees of depression at high levels of symptom severity.

Discussion

This study examined the process and accuracy of daily life-based lay assessments of subclinical depression. Such lay assessments are important for clinical research and practice because peers often serve as a critical link between the patient and mental health professionals. Three major findings emerged from the analyses: (1) Overall, across the entire sample, judges showed little accuracy at discriminating participants with low levels of depressive symptoms from those with high levels of depressive symptoms on the basis of ambient sounds sampled from 2 days of their lives. This finding is consistent with prior research on private personality traits such as neuroticism, or negative emotionality (John & Robins, 1993; Watson et al., 2000). (2) Judges’ lay assessments were, however, more accurate at high levels of depressive symptoms; among moderately and severely depressed participants, judges discriminated more reliably between degrees of depression. This finding is consistent with prior clinical research on deficits in social and emotional functioning in depression (Coyne et al., 1990; Rottenberg & Gotlib, 2004; Segrin, 2000). (3) Judges’ increased accuracy among moderately and severely depressed participants coincided with emerging behavioral manifestations (i.e., cue-validity correlations) at high levels of subclinical depression that were consistent with judges’ pattern of cue utilization.

Why were judges overall not more accurate in their lay assessments? Compared with other minimal-acquaintance scenarios that researchers have used for studying lay judgments of personality and clinical characteristics (e.g., Ambady & Rosenthal, 1992; Borkenau, & Liebeler, 1992; Funder & Sneed, 1993; Oltmanns, Friedman, Fiedler, & Turkheimer, 2004), the EAR sounds provided a wealth of information about participants’ naturally-occurring social as well as nonsocial episodes. Further, within the same paradigm (Mehl et al., in press), judges were not only quite accurate at rating participants on a highly observable trait such as Extraversion ($r = .41$), they also showed considerable accuracy in their ratings of more private traits such as Agreeableness ($r = .25$).
and Emotional Stability ($r = .30$). This suggests that judges’ lack of accuracy when rating levels of depression from the ambient sounds was not a result of insufficient information. Instead, the findings may hint at an interesting perceptual phenomenon.

The cue-utilization analyses suggest that depression was inferred from a lack of social engagement and pleasant activities: spending time alone, not talking (particularly in groups), not going out, not socializing, and not laughing were correlated with perceptions of depression. The correspondence between judges’ cue utilization and the emerging behavioral manifestations at high levels of symptomatology may indicate, then, that the cues judges used were derived from prototypes of highly depressed persons. When judges applied these cues undifferentiatedly across the entire spectrum of depression, they tended to be accurate in their assessments of truly depressed participants but inaccurate in identifying non- or only slightly depressed individuals.\(^5\)

The discontinuity in behavioral manifestations of depression, however, is likely a result of daily behaviors being causally overdetermined. People spend time alone, do not talk, or do not laugh for various reasons other than being depressed. Only at high levels of symptom severity does depression absorb sufficient variance in daily behaviors to lead to detectable behavioral manifestations that then can serve as cues for assessing depression. At low levels of depressive symptoms, these cues lack discrimination and render lay assessments of depression prone to bias, namely false positive judgments. Ironically, being falsely labeled as depressed may elicit social reactions that can ultimately foster the very development of depression (Coyne et al., 1990).

The paradigm used in this study deviates from naturalistic peer assessments in at least two important ways. First, assessments derived from the EAR-sampled ambient sounds lack critical visual cues such as facial expressions, gestures, or body postures. Providing a visual observer’s perspective may increase judges’ accuracy. Second, peers generally have access to baseline information about the targets’ social behaviors. Noticeable deviations from baseline may be particularly diagnostic cues for peer assessments of depression. Although this suggests that the accuracy estimates in this study may underestimate laypersons’ true accuracy, it is conceivable that a prototype-based assessment strategy would yield comparable degrees of accuracy and bias if visual or change information about targets’ daily social encounters were made available.

The study also raises important questions about differences between lay assessments and assessments by expert clinicians (Garb, 1998). To what extent would expert clinicians be more accurate than laypersons at identifying depressive symptoms from the ambient sounds? To what extent would expert clinicians be sensitive to or be similarly oblivious to the possibility that behavior considered prototypic of depression may not discriminate uniformly across all levels of symptomatology? An important direction for future research thus lies in directly comparing the diagnostic perspective of the self, the peers, and the clinician in a representative community sample. Such research has the potential to reveal perspective-specific diagnostic advantages and disadvantages and can thereby help increase the accuracy with which depression can be assessed where it is naturally encountered first—in people’s ordinary daily lives.

\(^5\) The effect of increased perceiver accuracy at high trait levels did not emerge for other traits. Perceiver accuracy was comparable among participants above ($r = .21$) and below ($r = .16$) the median in Extraversion as well as among participants above ($r = .12$) and below ($r = .21$) the median in Emotional Stability. Further, for both traits, perceiver accuracy was also not systematically higher in the highest (or lowest) tercile compared with the other two terciles of the distribution. This indicates that the effect may be rather specific to the assessment of depressive symptoms—perhaps because of the asymmetric distribution of the involved variables (Fisher, 1959).

References


Members of Underrepresented Groups: Reviewers for Journal Manuscripts Wanted

If you are interested in reviewing manuscripts for APA journals, the APA Publications and Communications Board would like to invite your participation. Manuscript reviewers are vital to the publications process. As a reviewer, you would gain valuable experience in publishing. The P&C Board is particularly interested in encouraging members of underrepresented groups to participate more in this process.

If you are interested in reviewing manuscripts, please write to the address below. Please note the following important points:

- To be selected as a reviewer, you must have published articles in peer-reviewed journals. The experience of publishing provides a reviewer with the basis for preparing a thorough, objective review.

- To be selected, it is critical to be a regular reader of the five to six empirical journals that are most central to the area or journal for which you would like to review. Current knowledge of recently published research provides a reviewer with the knowledge base to evaluate a new submission within the context of existing research.

- To select the appropriate reviewers for each manuscript, the editor needs detailed information. Please include with your letter your vita. In the letter, please identify which APA journal(s) you are interested in, and describe your area of expertise. Be as specific as possible. For example, “social psychology” is not sufficient—you would need to specify “social cognition” or “attitude change” as well.

- Reviewing a manuscript takes time (1–4 hours per manuscript reviewed). If you are selected to review a manuscript, be prepared to invest the necessary time to evaluate the manuscript thoroughly.

Write to Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.