

Classification of Domestic Cat (*Felis catus*) Vocalizations by Naive and Experienced Human Listeners

Nicholas Nicasastro and Michael J. Owren
Cornell University

To test for possible functional referentiality in a common domestic cat (*Felis catus*) vocalization, the authors conducted 2 experiments to examine whether human participants could classify meow sounds recorded from 12 different cats in 5 behavioral contexts. In Experiment 1, participants heard single calls, whereas in Experiment 2, bouts of calls were presented. In both cases, classification accuracy was significantly above chance, but modestly so. Accuracy for bouts exceeded that for single calls. Overall, participants performed better in classifying individual calls if they had lived with, interacted with, and had a general affinity for cats. These results provide little evidence of referentiality, suggesting instead that meows are nonspecific, somewhat negatively toned stimuli that attract attention from humans. With experience, human listeners can become more proficient at inferring positive-affect states from cat meows.

Animal communication has been variously described as a process of sharing information (Marler, 1984; Smith, 1997), of manipulation and skeptical reception (Dawkins & Krebs, 1978; Krebs & Dawkins, 1984), of “assessment/management” (Owings & Morton, 1998), and of “affect induction” in receivers (Owren & Rendall, 1997), among other approaches. The function of animal signals has also been diversely characterized as providing emotional readouts, fostering social prestige (Zahavi, 1975), or coding information in ways similar to human symbolic language (Cheney & Seyfarth, 1990; Macedonia & Evans, 1993).

Effective information sharing with, manipulation of, or management of humans is critically important to animals that depend on them. Encoding information would seem particularly relevant to communication in domestic animals. Although a number of wild species use calls that reportedly feature some degree of semanticity (for instance, vervet monkeys, Cheney & Seyfarth, 1990; yellow-bellied marmots, Blumstein & Armitage, 1997; and ravens, Bugnyar, Kijne, & Kotschal, 2001), the acquisition of functionally referential signals should be especially beneficial for domesticates, insofar as most cats and dogs live in environments where human referential signaling is ubiquitous.

There is evidence for unsuspected perceptual–cognitive abilities in domesticated species. Studies of social gaze in dogs have suggested canine capabilities of “reading” the posture, head position, and eyes of humans that compare favorably to those reported for chimpanzees (Hare & Tomasello, 1999) and even for 3-year-

old children (Soproni, Miklósi, Topál, & Csányi, 2001). Controlled experiments involving dog–owner play communication have concluded that dogs can competently interpret certain intentional movements of humans (McKinley & Sambrook, 2000; Rooney, Bradshaw, & Robinson, 2001). Feddersen-Petersen (2000) recently reported referential-like qualities in canine barks. There is also an extensive body of results attesting to the referential qualities of food and antipredator calls of domestic chickens (Evans & Evans, 1999; Gyger, Marler, & Pickert, 1987).

In their review of the problem of meaning in animal signals, Macedonia and Evans (1993) posited two criteria for functional referentiality: first, that calls be strongly correlated with a particular referent and second, that perceiver responses be correlated with the calls. Here, we describe two controlled experiments examining referentiality in domestic cat vocal signaling by testing whether humans can classify instances of a typical domestic cat-to-human vocalization in a manner consistent with the second of these criteria.

Moelk (1944) and Bradshaw and Cameron-Beaumont (2000) divided the vocal repertoire of the domestic cat into three broad categories: calls “produced with the mouth closed” (e.g., purrs and trills), “sounds produced while the mouth is held open in one position” (e.g., spitting or hissing), and calls produced “while the mouth is open and gradually closed” (Bradshaw & Cameron-Beaumont, 2000, p. 71). The last category constitutes the primary subject of this study: the *meow* or *miaow*. This call is characterized by a mean fundamental frequency (F_0) of ~400–1200 Hz, a modulating pitch profile, and a duration ranging from 110 to 3,100 ms in our sample (see Figure 1).

There are several reasons to pay particular attention to the use and evolution of this call. First, it is the most common cat-to-human vocalization (Bradshaw & Cameron-Beaumont, 2000); second, it is seldom observed in cat-to-cat interactions (Brown, 1993); and third, although undomesticated felids meow as juveniles, they rarely produce the call to humans in adulthood (Cameron-Beaumont, 1997). On the basis of spectrographic analysis, Shipley, Carterette, and Buchwald (1991) have described the meow as a periodic sound with F_0 corresponding to the vibration rate of the

Nicholas Nicasastro and Michael J. Owren, Department of Psychology, Cornell University.

Nicholas Nicasastro was supported by National Institute of Mental Health Predoctoral Research Fellowship T32 MN 19389 during preparation of this study. We thank cat owners Leslie Horowitz, Thomas Volman, Isabel Tovar, Jill Mateo, Darryl Mayeaux, S. K. List, Karl and Pamela Reichert, and Robert and Linda Newton. Thanks also to Penny Bernstein for helpful comments on the manuscript.

Correspondence concerning this article should be addressed to Nicholas Nicasastro, Department of Psychology, Cornell University, 243 Uris Hall, Ithaca, New York 14853. E-mail: nn12@cornell.edu

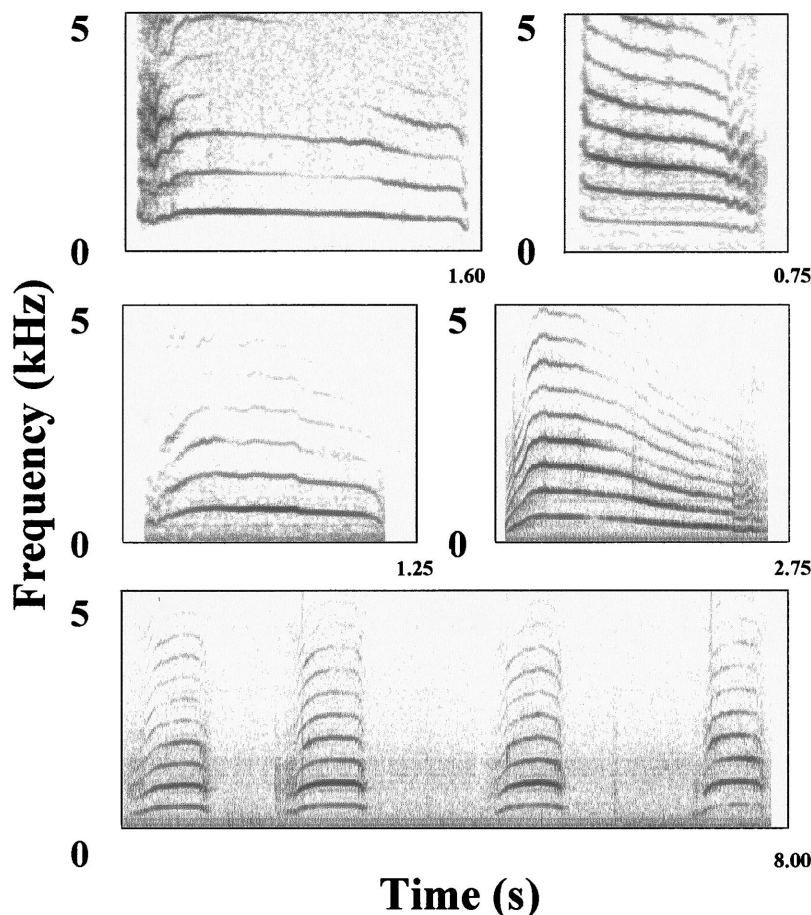


Figure 1. Narrow band spectrograms of four representative single meow calls and one bout. The contexts are as follows: top left = obstacle; top right = food related; middle left = distress; middle right = agonistic; and bottom = an affiliative bout. Spectrograms were generated by PRAAT sound analysis software (Boersma & Weenink, 2001); 22.05 kHz sampling frequency; window size = 0.029 ms; 90% overlap, no preemphasis.

vocal folds, a harmonic stack, and at least one band of tonal energy enhanced by the resonant properties of the vocal tract. The pitch profile is generally arched, with a rise in the center corresponding to maximum mouth opening. This is the most common, but not the only, pattern of articulation associated with the call. Other movements observed in the course of this study include little or no mouth opening, doubled openings, and coincidence of meowing and yawning. Although some studies of cat vocal behavior imply that the meow is fairly stereotyped, the call type very often includes atonal features (clicks, chaotic elements, etc.; Wilden, Herzel, Peters, & Tembrock, 1998) and garnishments (trilled leads or endings) that may serve as further bases of perceptual differentiation. Notwithstanding the categories set by Bradshaw and Cameron-Beaumont (2000), a significant number of the meows obtained in naturalistic contexts for our sample show such features. The wide degree of spectral and temporal variation in meows should provide the raw material on which selection (natural or developmental) may act. The meow is therefore a good candidate for a call that has been under selection pressure related to domestication.

How specific is the information listeners derive from cat-to-human vocal signals? Do the graded acoustic properties of cat

meows, despite broad individual and situational variation, render it possible for humans to glean discrete, quasi-referential messages from these signals, like those reported for antipredator alarm calls in several species of wild primates, sciurids, and birds (Blumstein & Armitage, 1997; Evans & Evans, 1999; Struhsaker, 1967)? With experience, can humans learn to discriminate between the calls? Alternately, would inaccurate responding to cat meows by human listeners at least show consistency, including consistent patterns of errors? The larger goal of these and future studies is to develop a better understanding of cat-human vocal interaction.

This article describes the results of two laboratory studies designed to answer these questions by testing human perceptual responses to cat meows. These vocalizations were recorded in five different, well-defined contexts in which cats and humans commonly interact. In Experiment 1, participants with various degrees of experience with cats attempted to classify single meows by behavioral context. In Experiment 2, the classification task was performed on bouts of meows.

Experiment 1

Human listeners were presented with recordings of single cat meows and asked to classify them by context of production. The

production contexts were chosen for familiarity and clear correlation with associated cat behaviors (see Table 1). Assignment of calls to particular contexts here is intended to mean only that the calls were closely associated with nonvocal behaviors typical of the contexts involved. It is not meant to suggest any knowledge of the internal states or intentions of the animals on the part of the experimenters.

Method

Participants. Participants were 33 adult human listeners with varying familiarity with cats in general, but all were unfamiliar with the subject cats. Data from 5 participants were dropped because of experimenter error that affected the stimuli they heard. The remaining 28 participants included 15 women and 13 men, with a mean age of 23.90 years ($SD = 7.98$). Nineteen college-age participants took part in the experiments for extra credit in courses, and 9 others received a food (chocolate) reward.

Materials. A corpus of 535 cat meows was recorded in the home environments of 12 adult cats (7 females and 5 males), including 2 sibling animals owned by Nicholas Nicastro. Other owners were all acquaintances or relatives of Nicholas Nicastro, and none of their cats were related. Cats were all adults, but because some had been adopted from shelters as adults, precise ages were unknown for 6 of the animals. Of the remaining 6, mean age was 6.83 years ($SD = 3.43$), range 2–13 years.

Calls from the various contexts were recorded by Nicholas Nicastro on an ad libitum basis, with the microphone placed at the smallest distance from the cats that did not noticeably alter their behavior (~0.5–2.0 m). Owners were asked to confirm that their cats were behaving as usual during the recording sessions. In virtually all cases, after the first hour or so of recording, the owners reported their cats were behaving typically. Variable microphone distance necessarily creates an additional source of variability in amplitude of the recordings. We compensated for this variation in test stimuli preparation using the method described below.

Meows were recorded from five behavioral contexts, including *food related* (calls made prior to regular feeding), *agonistic* (calls made when the cats were antagonized by their owners, such as during vigorous brushing), *affiliative* (calls made when the cat solicited affection from their owners), *obstacle* (calls made when cats solicited help to negotiate a barrier, such as a closed door or window), and *distress* (calls made when the cats were placed in an unfamiliar environment—in all cases here, when the animals were placed in automobiles). The behavioral definitions of the calling contexts are presented in Table 1. All animals gave at least some

single calls. Eight animals vocalized both in single calls and in bouts, often within the same recording session.

Recordings were made with a Sennheiser ME-88/K3U directional microphone (Sennheiser GmbH, Wedemark, Germany) and either a UHER 4200 Report Monitor (Uher Werke Munchen, Munich, Germany) or a Marantz PMD222 portable cassette recorder (Superscope Technologies, Aurora, IL). Calls were digitized at a 22.05 kHz sampling rate and examined and edited on a Silicon Graphics O2 workstation (Silicon Graphics, Inc., Mountain View, CA) using ESPS/waves+ 5.1 acoustic-analysis software (Entropic Research, Washington, DC). Twenty calls were selected per production context for a total of 100 per experimental block. Approximately 100 of the 535 total meows were eliminated outright because of excessive background noise or the presence of telltale indications of behavioral context. Meows clearly made during bouts were held back for Experiment 2. Of the approximately 150 remaining single calls, the 100 recordings with the least residual background noise (including noncontextual-specific cues such as animal-related, recordist, and machine noises) were selected for the experiment. Other than clarity and quality of the recording, no other acoustic criteria were applied in the selection of the test meows.

Descriptive statistics for basic acoustic properties of the calls are presented in Table 2. Generally speaking, our sample shows properties much like those reported for meows elsewhere (e.g., Bradshaw & Cameron-Beaumont, 2000). With only a few exceptions (see *Results and Discussion* of Experiment 1), calls from the five contexts were not acoustically distinct but instead appear to lie on graded continua of durations and pitches.

As some cats did not vocalize in some situations, not all the animals were represented by calls in each context. Nor was each animal represented by equal numbers of calls or bouts. However, at least 5 different cats were included in each sample of 20 meows representing each call context ($M = 5.80$ different cats per context, $SD = 0.84$).

Meows were prepared as stimuli using custom sound format conversion software (Tice & Carrell, 2002), SpeechStation II (Sensimetrics, Cambridge, MA), and the ESPS/waves+ program. Stimuli began and ended with 100-ms silent segments, which helped prevent audible transients during presentation. To roughly equate the calls for loudness, we normalized stimuli by rescaling each waveform so its highest amplitude peak was set to the maximum representable value. During presentation, stimuli were low-pass filtered at 10 kHz, mixed with a low-amplitude white noise, and attenuated to a comfortable listening level. The background noise ran continuously throughout each experimental session.

Design and procedure. Participants were tested in groups of 1 to 3 individuals in a room with five individual listening booths. Testing was controlled from a computer in an adjacent room, which operated Tucker-Davis Technologies (TDT; Gainesville, FL) System II audio modules and recorded the responses. In a randomized block, repeated measures design, participants were asked to classify the 100 stimuli by context. The calls were presented once each, one at a time, in random order, via beyerdynamic DT 831 headphones (beyerdynamic North America, Farmingdale, NY). Each trial consisted of a 1,250-ms cueing interval, a maximum of 6 s in which to make a context selection, and a 1-s intertrial interval. Responses were made on the middle five buttons of seven-button TDT response boxes. The linear order of context labels on the response boxes was alternated and counterbalanced across participants.

The participants were told that “the purpose of this experiment is to understand what humans know about the vocalizations of cats.” Background data were collected before listening trials using a questionnaire. Here, the participants used a 7-point scale to rate their experience and affinity with cats in three ways: how recently they had lived with a cat (residency: 1 = *never lived with a cat* to 7 = *currently live with a cat*), their estimated rate of interaction with all cats (interaction: 1 = *never interact with cats* to 7 = *interact several times a day*), and their affinity for cats in general (affinity: 1 = *very negative* to 7 = *very positive*). After testing, the questionnaire was returned and the participants were asked to make a single additional judgment: to rate the difficulty of the classification task

Table 1
Behavioral Definitions of Vocal Production Contexts

Context variable	Definition
Food related	Cat orients toward owner and/or food; eyes open, looking alternately from owner to food; ears up; tail up.
Agonistic	Cat orients toward offending object (e.g., hand or hair-brush); spitting; growling; ears back; eyes wide open; tail down and mobile.
Affiliative	Cat orients toward owner; leg rubbing; purring; eyes closed or half open; tail up.
Obstacle	Cat orients toward obstacle (e.g., window or door); eyes open, looking alternately from owner to obstacle; tail up.
Distress	No specific orientation, with repetitive movement such as pacing; eyes wide open; ears up but slightly back; tail down and mobile.

Note. Calls were assigned to contexts only if the listed behaviors were observed.

Table 2
*Statistics for Basic Acoustic Measures of 100 Cat Meows
 Presented in Experiment 1*

Acoustic measure	<i>M</i>	<i>SD</i>	Range	Context correlation ^a
Duration (s)	0.78	0.45	0.11–3.10	(Agonistic +)**
Mean F_0 (Hz)	621	165	372–1039	(Agonistic –)**
Initial F_0 (Hz)	581	233	172–1119	(Agonistic –)*
Final F_0 (Hz)	592	211	190–1110	(Agonistic –)*
Max F_0 (Hz)	767	191	431–1335	<i>ns</i>

Note. F_0 = fundamental frequency.

^a Context correlation is the significance level for one-way analysis of variance for main effect of production context on the acoustic measure. $\alpha = .05$. Contexts in parentheses are those most distinctly differentiated from the others by post hoc analysis (Tukey's pairwise comparisons); the direction of differentiation (+ = higher, – = lower) is also indicated.

* $p < .05$. ** $p < .01$.

(1 = *very easy* to 7 = *very difficult*). The overall testing sessions took 25–35 min. Potential bias in classification of calls presented later in the sessions because of fatigue was controlled by randomizing the order of presentation.

Data analysis. The listeners' accuracy in classifying cat meows was evaluated in two ways, by individual context and by grouping according to positive and negative affective valence. A context accuracy ratio (C^R) was calculated by dividing overall observed hit rates by the rate expected by chance (20%). $C^R > 1$ thus denotes better-than-random performance, whereas $C^R = 1$ denotes chance responding. In addition to normalizing performance relative to chance, this measure was found consistently to produce near-normal response distributions, obviating the need for other transformations.

In grouping the contexts by their presumptive affective valence, we assigned food-related and affiliative to the positive affect category, whereas agonistic and distress were considered to be negative. Obstacle responses were not included in this analysis because that context is more difficult to characterize unambiguously as either positive or negative. Participants' percentage accuracy in classifying the resulting 80 calls by positive versus negative contexts was then reduced to an affect accuracy ratio (A^R) of observed performance versus chance (50%).

Statistical analyses included standard and repeated measures analysis of variance (ANOVA), one- and two-sample *t* tests, and simple regression. Post hoc analysis was conducted using the Tukey–Kramer method of pairwise comparisons and Bonferroni (all pairwise) multiple comparison tests. Analyses were implemented using MINITAB 12.2 (Minitab Inc., State College, PA) and NCSS statistical software (Jerry Hintze, Kaysville, UT). An alpha value of .05 was used for all statistical tests.

Results and Discussion

A repeated measures ANOVA showed no significant main effects on context classification accuracy for participant age, $F(12, 15) = 2.13$, *ns*, or participant judgment of difficulty of the experimental task ($M = 3.29$, $SD = 1.76$), $F(5, 22) = 1.56$, *ns*. Participant sex approached but did not reach significance, $F(1, 26) = 4.14$, $p > .05$, and no significant interactions occurred among these factors. Classification of affective valence showed no main effects for participant age, $F(12, 15) = 1.38$, *ns*; participant sex, $F(1, 26) = 2.36$; *ns*; judgment of difficulty, $F(5, 22) = 0.77$, *ns*; or any of their interactions.

The pooled mean performance ratio C^R for the context classification task was 1.36 ($SD = 0.30$), or 27.2% correct responses. This performance was only modestly accurate but significantly above chance, $t(27) = 6.57$, $p < .01$. No participant performed at

less than an accuracy ratio of .85. There were significant main effects for affinity for cats, $F(5, 22) = 3.74$, $p = .01$; for residency with cats, $F(4, 23) = 5.79$, $p < .01$; and for frequency of interaction with cats, $F(5, 22) = 3.59$, $p = .02$, as illustrated in Figure 2. No significant interactions were found. Regression analysis showed that the effect sizes for all these measures of familiarity were substantial: Residency, interaction, and affinity accounted for 35.5%, 40.2%, and 43.1% of the variance, respectively. Post hoc analyses indicated that more recent residency, higher frequency of interactions, and greater affinity for cats were all associated with greater accuracy in the context-classification task.

Results of repeated measures ANOVAs for the effect of the experience and affinity measures on classification accuracy for the five individual production contexts are shown in Table 3. In summary, affiliative and agonistic were the only contexts that showed significant main effects for at least two of the experience measures. Two of these effects retained significance after applying the Bonferroni correction for multiple tests (Darlington, 1990). It is interesting to note that post hoc analyses of the effects of call context on the acoustic measures showed that agonistic was the most acoustically distinct call. Specifically, agonistic calls were longer in duration and lower in overall mean, mean initial, and mean final F_0 than calls produced in the other contexts.

For classification of affective valence, the pooled performance ratio A^R for the 28 participants was 1.07 ($SD = 0.17$), or 53.7% accuracy. This performance is slightly but also significantly above chance (50%), $t(27) = 2.38$, $p = .03$. Overall, classification of the affective valence of the calls showed a less marked effect of participant experience. An ANOVA showed significant main effects on classification accuracy of residency with cats, $F(5, 22) = 5.11$, $p < .01$, $R^2 = .253$, and frequency of interaction, $F(5, 22) = 2.66$, $p = .05$, $R^2 = .266$, but not for affinity for cats, $F(5, 22) = 2.05$, *ns*. Examining classification accuracy for positive and negative affect separately, an ANOVA revealed only one significant result for the experience measures: Participants reporting more recent residency with cats showed an improvement in their ability to correctly classify production contexts with positive affective valence, $F(4, 23) = 4.77$, $p < .01$, $R^2 = .242$. Classification of negatively toned calls did not significantly improve on the basis of any of the experience and affinity measures.

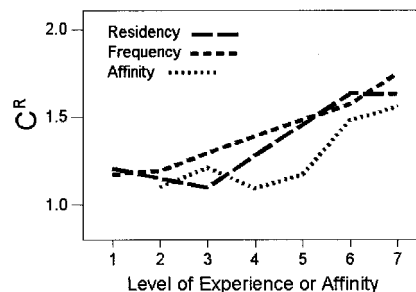


Figure 2. The effect of two experience measures (residency and frequency) plus affinity on mean classification accuracy measured relative to chance performance (C^R) for single meows. For clarity, category means are connected as continuous functions, each of which showed a statistically significant effect of experience and affinity on accuracy. No participants reported Level 1 (lowest) affinity for cats. $C^R = 1$ denotes performance at chance level.

Table 3
Significance Levels for Multiple One-Way Analysis of Variance for Main Effects of Experience Measures on Classification Accuracy for Each Production Context, $\alpha = .05$

Context	Residency	Interaction	Affinity
Single calls			
Food related	<i>ns</i>	<i>ns</i>	<i>ns</i>
Affiliative	** (*)	* (<i>ns</i>)	.07 (<i>ns</i>)
Agonistic	** (<i>ns</i>)	** (**)	* (<i>ns</i>)
Obstacle	<i>ns</i>	<i>ns</i>	** (*)
Distress	<i>ns</i>	<i>ns</i>	<i>ns</i>
Bouts			
Food related	* (<i>ns</i>)	<i>ns</i>	<i>ns</i>
Affiliative	<i>ns</i>	<i>ns</i>	<i>ns</i>
Agonistic	<i>ns</i>	<i>ns</i>	<i>ns</i>
Obstacle	<i>ns</i>	<i>ns</i>	<i>ns</i>
Distress	<i>ns</i>	<i>ns</i>	<i>ns</i>

Note. Symbols in parentheses are Bonferroni-corrected significance levels. Levels that just miss significance are listed numerically.

* $p < .05$. ** $p < .01$.

To check for the possibility that our results were an artifact of possible ambiguity in some of the contexts selected for the study, we analyzed participant performance in classifying the contexts deemed most straightforward: food related, affiliative, and distress. (Agonistic was eliminated from this phase of the analysis because it is arguably characterized better by other call types; obstacle was left out because it suffers from potential affective ambiguity.) Mean C^R fell from 1.36 with all five contexts to 1.18 with three. This difference missed statistical significance, $t(48) = 1.88$, *ns*, but the adjusted C^R was still significantly above chance performance (one-sample, two-tailed test, $t[25] = 2.51$, $p = .02$). Regression analyses of the effect of experience and affinity on single call classification accuracy showed significant effects for frequency of interaction, $F(1, 23) = 5.27$, $p = .03$, and affinity, $F(1, 23) = 5.27$, $p = .03$, but not for residency, $F(1, 23) = 3.35$, *ns*. In other words, although there were some differences, the reanalysis based on three contexts showed the same general pattern of results as with five. The overall robustness of these results suggests, therefore, that they had not been substantively influenced by ambiguities in the context definitions.

Our results suggest that humans can, with experience, improve their reading of the production context and, to a lesser degree, the affective valence of cat meows. Hit rates were modest in both cases, however: The most accurate participant on the context classification task was a 30-year-old woman who reported the highest level of affinity for cats, the highest frequency of interaction with cats, and living with a cat within the last 6 months. She achieved an accuracy ratio of 2.05, or 41% correct. Compared with, say, hit rates of approximately 90% or more reported for humans classifying human vowel sounds, this is not very accurate (cf. Hillenbrand & Nearey, 1999).

In addition to showing only modest accuracy, participants in Experiment 1 were also inconsistent in their responses. In other words, there was little indication that when listeners were incorrect their errors were systematically patterned. For 80 of 100 stimuli, no single response category was chosen by 50% or more of the participants. Of the 20 stimuli for which one response choice did garner 50% or more of the responses, 10 were agonistic and 7 were

affiliative. This distribution is not surprising, considering that these were exactly the contexts that showed evidence for classification accuracy improvement with experience or affinity.

We also examined whether classification accuracy had been affected by biased responding. Participants could have achieved elevated accuracy rates for some contexts by, for instance, attributing calls to those contexts more often. This strategy could not have improved the aggregate accuracy rate across all contexts, however, because additional hits due to bias would necessarily be offset by additional misses for the others. To check that bias was not responsible for higher accuracy rates for any one context, we used standard signal-detection procedures to derive a bias measure (c) for each participant's classification of agonistic, the context with the highest classification accuracy across participants (Macmillan & Creelman, 1991). Regression analysis revealed a highly negative correlation between degree of bias and accuracy, $F(1, 26) = 122.70$, $p < .01$, $R^2 = .842$. In other words, participants who performed best at the classification task were ones who showed values of c closest to the midpoint 0, meaning minimal bias. It therefore follows that systematic bias did not increase classification accuracy with respect to this context.

We also found a negative correlation between degree of bias and one of the measures of experience with cats. Regression analysis for residency on bias c showed a negatively sloped function, with $F(1, 26) = 4.53$, $p = .04$, $R^2 = .165$.

Experiment 2

In this experiment, we tested classification accuracy for context and affective valence for the next higher level of complexity in cat vocalizations, namely bouts of meows. A bout was defined as a sequential series of at least two meows given within a single behavioral context and in close temporal proximity to each other (Martin & Bateson, 1986; see Figure 1). The production contexts were identical to those in Experiment 1.

Method

Participants. To control for the effect of familiarity with the task, we recruited a different sample of 28 adult participants (15 women, 13 men; mean age = 27.04 years, $SD = 15.60$). Participants had varying familiarity with cats in general but were again unfamiliar with the subject cats. Twenty-one college-age participants took part in the experiments for extra credit in courses and 7 others received a food (chocolate) reward. Data from 2 participants were dropped from this experiment: one because of equipment problems and another because of personal familiarity with two of the cats recorded. Two-sample t tests comparing these 26 participants with the 28 in Experiment 1 on all experience and affinity measures showed no significant difference in chronological age, $t(52) = 0.95$, *ns*; length of residency with cats, $t(52) = -0.81$, *ns*; frequency of interaction with cats, $t(52) = 0.44$, *ns*; or affinity for cats, $t(52) = -0.13$, *ns*.

Materials and procedure. Recording specifications and test setup were the same as for Experiment 1, with the following exception: Instead of 100 individual meows, participants heard a total of 40 bouts composed of between 2 and 6 meows ($M = 3.30$, $SD = 0.85$). In 15% of the candidate bouts, intermittent instances of excessive background noise obliged us to edit out either the first or the last meow in the bout. In no case was the temporal order of the meows or the intercall intervals altered. Data analyses were the same as for Experiment 1.

Eight of the 12 cats in our sample vocalized in bouts. As some cats did not vocalize in some situations, not all animals were represented in the context categories, nor were all represented by equal numbers of bouts.

However, at least 3 different cats were included in each sample of 8 bouts representing each vocalizing context ($M = 4.20$ different cats per context, $SD = 0.84$).

Results and Discussion

A repeated measures ANOVA showed no significant main effects on context classification accuracy for participant age, $F(11, 14) = 0.64$, *ns*; participant sex, $F(1, 24) = 0.10$, *ns*; participant judgment of difficulty of the experimental task, $F(5, 20) = 0.61$, *ns*; or any interactions of these factors. Likewise, there were no significant main effects on affect classification for participant age, $F(11, 14) = 0.66$, *ns*; participant sex, $F(1, 24) = 1.13$, *ns*; participant judgment of difficulty of the experimental task ($M = 2.80$, $SD = 1.50$), $F(5, 20) = 1.68$, *ns*; or any interactions of these factors.

Pooled mean C^R for the 26 participants was 1.72 ($SD = 0.47$), or 34.4% correct. This performance was significantly above chance, $t(25) = 7.18$, $p < .01$, and also significantly above the accuracy ratio for Experiment 1, $t(52) = 3.51$, $p < .01$. On affect classification, the pooled mean performance ratio was 1.09 ($SD = 0.22$), or 54.5% accuracy. This outcome was significantly above chance, $t(25) = 2.08$, $p = .05$, but not significantly better than A^R for Experiment 1, $t(52) = 0.27$, *ns*.

As illustrated in Figure 3 and in contrast to Experiment 1, there were no significant main effects on C^R for any of the experience measures, such as length of residency with cats, $F(5, 20) = 1.41$, *ns*, or frequency of interaction with cats, $F(6, 19) = 0.93$, *ns*, nor were there any significant interactions. Affinity for cats likewise showed no significant relation with accuracy, $F(5, 20) = 1.55$, *ns*, although plotting the outcomes shows a clearly ascending function. Unfortunately, too few participants reported affinity levels of 1 or 2 to afford a definite statistical conclusion with respect to this measure. Results still missed significance, however, when the few lowest affinity respondents were removed from the data set, leaving a uniform distribution of respondents across the remaining categories, $F(3, 20) = 1.96$, *ns*.

An ANOVA based on hit rates on individual contexts likewise showed no effect of the experience and affinity measures, except for residency in the case of food related, $F(5, 20) = 3.19$, $p = .03$

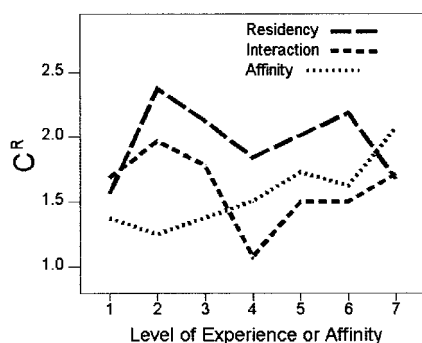


Figure 3. The effect of two experience measures (residency and interaction) plus affinity on classification accuracy measured relative to chance performance (C^R) for bouts. For clarity, category means are connected as continuous functions. None of the functions show a significant experience or affinity effect, although there appears to be a steady rise in accuracy with affinity (see text for discussion). $C^R = 1$ denotes performance at chance level.

(see Table 2). However, that p value rose to .42 (*ns*) when Bonferroni corrected for multiple tests and may therefore be attributable to chance. Classification of the affective valence of the bouts also failed to improve with greater experience. There were no significant main effects on classification accuracy of participant residency with cats, $F(5, 20) = 1.82$, *ns*, or frequency of interaction, $F(6, 19) = 1.04$, *ns*. Participant affinity for cats approached but did not reach significance, $F(5, 20) = 2.41$, $p = .07$.

Examining classification accuracy for positive and negative affect separately, we did find a significant effect of experience and affinity. Accuracy in classifying bouts produced in positively toned contexts did improve with more recent residency with cats, $F(5, 20) = 2.95$, $p = .04$, $R^2 = .173$, and with increased affinity for cats, $F(5, 20) = 3.23$, $p = .03$, $R^2 = .234$, but not with increased frequency of interaction, $F(6, 19) = 1.38$, *ns*. Classification of negatively toned bouts failed to improve significantly on the basis of any of the experience or affinity measures.

Checking again for the possibility that our results were an artifact of the five contexts selected, we analyzed participant performance in classifying food related, affiliative, and distress. Mean C^R fell from 1.72 with five contexts to 1.55 with three. This difference missed statistical significance, $t(50) = 1.06$, *ns*, and the adjusted C^R was still significantly above chance performance (one-sample, two-tailed test, $t[26] = 4.49$, $p < .01$). Regression analyses of the effect of experience and affinity on bout classification accuracy showed no significant outcomes for interaction, $F(1, 24) = 0.25$, *ns*, for residency, $F(1, 24) = 0$, *ns*, which was in accord with the results based on all five contexts. Affinity did show a significant effect, $F(1, 24) = 5.39$, $p = .03$, but only a few participants reported category 1 or 2 (lowest) affinity for cats, which clearly skewed the results at the low end. Rerunning the regression with the lowest respondents removed yielded a nonsignificant result as well, $F(1, 22) = 2.65$, *ns*. Although there were some differences, reanalysis based on three contexts showed the same general pattern of results as with five.

As in Experiment 1, participant responding in Experiment 2 showed neither great accuracy nor systematic patterning of errors. For 27 of the 40 stimuli, no response category was chosen by a simple majority of the participants. Of the remaining 13 stimuli, 5 were agonistic, 3 affiliative, 2 obstacle, 2 distress, and 1 food related. The slightly higher degree of consistency in Experiment 2 may be attributed to the slight but significant increase in classification accuracy. Once again, multiple post hoc comparisons for all stimuli showed all contexts to be regularly confused with at least several of the others.

Participants showed significantly better ability to classify production context in Experiment 2 than in Experiment 1. These results are not surprising insofar as listeners in Experiment 2 were presented with more total meows (132 vs. 100) and more potential cues than in Experiment 1, such as temporal patterning of calls. The additional information did not significantly improve classification accuracy for aggregate (positive and negative) affective valence, however.

Unlike with single calls, though, context-classification accuracy for bouts did not appear to significantly improve with participants' experience with cats. In Experiment 2, the most accurate participant was an 18-year-old woman who reported residency with a cat between 5 years and 1 year before testing and almost no interaction since but the highest level of affinity for cats in general. Her hit rate was 50%, exceeding chance performance by a factor of 2.5.

General Discussion

These results show that although humans can exceed chance performance in their discrimination among production contexts of single meows and bouts, they achieve only modest rates of accuracy. The most specific outcomes may be summarized as follows: (a) Participants classified bouts more accurately than single calls; (b) there was a facilitating effect of experience with cats on context classification accuracy for single calls, but not for bouts; and (c) there was either a small (single calls) or no (bouts) facilitating effect of experience and affinity on classifying affective valence.

We find these results to be inconsistent with Macedonia and Evans' (1993) second criterion for functional referentiality, that receiver response be reliably correlated with call type. Although cats clearly use these calls to communicate with humans, we have found little evidence of functional referentiality. Despite the potential advantage to cats of conveying specific messages to their owners, humans either do not derive specific information from the calls or achieve only imperfect classification skill based on experience.

The observed experience and affinity effect on classification of production context of single calls is consonant with this conclusion. Instead of conveying specific messages that would evoke a "get it" or "don't get it" response in listeners, sensitivity to production context seems to increase gradually with experience and affinity, and then only in agonistic or affiliative contexts. With respect to agonistic contexts, one concomitant change in experienced listeners was a decrease in bias.

In evaluating these results, it is important to bear in mind that the behavioral contexts used were chosen specifically because they are very common ones for cats and humans. However, it is conceivable that meows might make reference to only a subset of the contexts or to contexts we have not included. Classifying the calls of unfamiliar animals solely on the basis of disembodied sounds, without the associated physical cues (facial, tail, and body expression, orientation, etc.), is difficult. The variability of meows, both within and between contexts and individuals, further complicates the classification task in both experimental and natural circumstances. Rather than assert some absolute yet debatable criterion for what constitutes good performance, we might instead evaluate the results in a relative sense. One way to do this is to express the change in accuracy rates in terms of *percent error reduction*, meaning the degree to which participant error rate improved on that expected by chance (calculated as: $\{[(100 - \text{chance rate}) - (100 - \text{observed rate})] \times 100\} / [100 - \text{chance rate}]$). In this case, the 4 participants with the highest frequency of interaction (more than once a day) reduced their incorrect responses in single-call context classification by 15% over the 5 participants with the lowest frequency (never interact).

Accuracy rates for classifying the affective valence of single calls and bouts were even more modest than for context. Performance significantly exceeded chance, but barely. The effect of experience on classifying single meows was slight: Percent error reduction from the lowest to the highest frequency of interaction was only 6.25%. These results suggest that meows are not very good readouts of the cats' affective state at the moment of production. In postsession interviews, many participants reported that all the calls sounded negative, even though 40% were recorded in affiliative or food-related contexts.

Experiment 2 presented both expected and unexpected outcomes. On the basis of means for all levels of experience, listening to bouts instead of single calls decreased overall context classification error by 10%. Experienced participants in both experiments more accurately classified stimuli with positive affective valence, suggesting that they were more sensitive to vocalizations produced in positive circumstances. Yet with respect to context and aggregate affective valence, our participants showed no significant error reduction with experience.

The difference between the effects of experience in Experiments 1 and 2 was unanticipated. One possible explanation is that listeners used different strategies for classifying calls versus bouts and that only one of these strategies is correlated with experience. We saw no evidence for any conscious difference in listening strategy on the basis of postsession interviews. Presumably, then, any strategy difference would have been beneath awareness on a fairly low level of perceptual processing.

Whatever the reason for these differing outcomes, the pattern of listener responses to single calls versus bouts presents cats with an evident trade-off when attempting to inform or manage humans. Single calls are energetically cheap signals that may not immediately provoke specific responses but may become more effective if the cat and owner are of long acquaintance. By contrast, producing bouts of calls is marginally more energetically expensive but tends to provoke more consistent responses across all listeners. We might speculate, therefore, that cat domestication by humans selected for a greater tendency to produce bouts over single meows—in other words, for housecats to become more vocal than their wild forebears.

These studies do not replicate Feddersen-Petersen's (2000) observations of functional reference in dog barks. One explanation might be that referential signaling is relatively rare among nonhumans, despite the evident advantage it may afford vocalizers (Rendall & Owren, 2001). Our failure to show referential signaling might also be a consequence of the fact that cats, unlike dogs, are relatively recent domesticates, with the first record of cat-human cohabitation postdating large-scale agriculture in ancient Egypt (about 6,000 years ago; Serpell, 2000). Even after domestication, cats may have been relatively independent of their human caretakers for some needs. By contrast, dogs were almost certainly domesticated far earlier, in Neolithic times at the latest (Dayan, 1994; Vilá et al., 1997). Perhaps, therefore, either selection pressure was initially too weak or time has been too short, or both, for cats to evolve functionally referential signaling.

Because the meow represents a class of sounds and occurs in a variety of circumstances, other calls may be more context specific and hence more referential in function. For instance, listeners may have been much better able to label pure purrs as occurring in affiliative contexts or spits in agonistic ones. In fact, meows may be under reduced or no selective pressure toward functional referentiality in contexts in which other calls are available to be used. Nevertheless, cats do meow regularly to humans in the contexts we tested, so some degree of anthropogenic selection pressure should apply to the calls in all the contexts.

It is plausible that with repeated instances of interaction over developmental time cats can learn to incorporate acoustic cues that have specific referential value to their individual owners and that owners can better learn to attach meaning to the acoustic qualities of calls produced in specific contexts. Such conventionalization or "ontogenetic ritualization" (Tommasello & Call, 1997) would be

rooted in human caregivers' evident penchant for ascribing significance and even symbolic value to sounds and in cats' ability to modify meows in accordance with their effectiveness in eliciting desirable human response in particular circumstances. This process would not be reflected in tests of nonowners of unfamiliar cats.

If demonstrated, the occurrence of ontogenetic ritualization in communication between domestic cats and humans might help explain the modest classification ability of the nonowners tested here. Ritualization is predicated on the regular association of particular vocal cues with specific contexts or actions known to both the cat and its owner. Such cues are necessarily drawn from a repertoire of vocal behaviors that are to some degree shared by all cats. Experienced nonowners might conceivably classify the calls of unknown cats on the basis of apparent resemblances with the context-specific calls of known ones. However, given the inherent variability of meows in general, differences in temperament and propensity to vocalize between cats, and the challenge of classifying context-free, disembodied calls, analogies with known cats would presumably be of only limited help. The outcome would be something like the highly imperfect classification skill we observed among experienced listeners.

Fundamentally, we suggest that the primary selection pressure operating on cat meows has been toward producing sounds that better exploit typical human auditory and affective responses. In other words, the domestication process is likely to have prompted more meowing and using meows that are more salient to humans in particular. The pitch-varying, tonal structure of meows may be telling in this regard, as sensitivity to frequency modulation is significantly greater in humans than in cats or in other mammals tested (reviewed in Fay, 1992). The primary effect of these salient, somewhat negative-sounding calls may thus be to provoke, rather than to specify, a reaction in humans.

On the basis of our results, it appears that human caregivers can become more attuned to cues to cats' positive affect implicit in the calls. Contrasts in affective valence (positive vs. negative) might then create the foundation for a "bootstrapping" ritualization process whereby owners become better attuned to their cats' emotional states and specific needs in particular contexts. This hypothesis is being examined in ongoing studies; for instance, by measuring the degree to which owners themselves are able to infer vocalizing contexts from recorded meows of their own cats. Other work is examining the basis of the modest success that has been demonstrated among nonowners by examining how human affective responses to the acoustic properties of these sounds may be providing some guidance to inferring cats' emotional states. There is still much to be learned about this complex, but instructive, communicative relationship.

References

- Blumstein, D. T., & Armitage, K. B. (1997). Alarm calling in yellow-bellied marmots: I. The meaning of situationally variable alarm calls. *Animal Behaviour*, *53*, 143–171.
- Boersma, P., & Weenink, D. (2001). Praat (Version 4) [Computer software and manual]. Retrieved from <http://www.fon.hum.uva.nl/praat/>
- Bradshaw, J. W. S., & Cameron-Beaumont, C. L. (2000). The signaling repertoire of the domestic cat and its undomesticated relatives. In D. C. Turner & P. Bateson (Eds.), *The domestic cat: The biology of its behaviour* (2nd ed., pp. 67–93). Cambridge, England: Cambridge University.
- Brown, S. L. (1993). *The social behaviour of neutered domestic cats*. Unpublished doctoral dissertation, University of Southampton, England.
- Bugnyar, T., Kijne, M., & Kotschal, K. (2001). Food calling in ravens: Are yells referential signals? *Animal Behaviour*, *61*, 949–958.
- Cameron-Beaumont, C. L. (1997). *Visual and tactile communication in the domestic cat (Felis silvestris catus) and undomesticated small felids*. Unpublished doctoral dissertation, University of Southampton, England.
- Cheney, D. L., & Seyfarth, R. M. (1990). *How monkeys see the world*. Chicago: University of Chicago.
- Darlington, R. (1990). *Regression and linear models*. New York: McGraw-Hill.
- Dawkins, R., & Krebs, J. R. (1978). Animal signals: Information or manipulation? In J. R. Krebs & N. B. Davies (Eds.), *Behavioural ecology: An evolutionary approach* (pp. 282–309). London: Blackwell.
- Dayan, T. (1994). Early domesticated dogs of the Near East. *Journal of Archaeological Science*, *21*, 633–640.
- Evans, C. S., & Evans, L. (1999). Chicken food calls are functionally referential. *Animal Behaviour*, *58*, 307–319.
- Fay, R. R. (1992). Structure and function in sound discrimination among vertebrates. In D. B. Webster, R. R. Fay, & A. N. Popper (Eds.), *The evolutionary biology of hearing* (pp. 229–263). New York: Springer-Verlag.
- Feddersen-Petersen, D. (2000). Vocalization of European wolves (*Canis lupus lupus L.*) and various dog breeds (*Canis lupus f. fam.*). *Archives of Animal Breeding*, *43*, 387–397.
- Gyger, M., Marler, P., & Pickert, R. (1987). Semantics of an avian alarm call system: The male domestic fowl, *G. domesticus*. *Animal Behaviour*, *102*, 15–40.
- Hare, B., & Tomasello, M. (1999). Domestic dogs (*Canis familiaris*) use human and conspecific social cues to locate hidden food. *Journal of Comparative Psychology*, *113*, 173–177.
- Hillenbrand, J. M., & Nearey, T. M. (1999). Identification of resynthesized /hVd/ utterances: Effects of formant contour. *Journal of the Acoustical Society of America*, *105*, 3509–3523.
- Krebs, J. R., & Dawkins, R. (1984). Animal signals: Mind-reading and manipulation. In J. R. Krebs & N. B. Davies (Eds.), *Behavioral ecology: An evolutionary approach* (2nd ed., pp. 380–402). Oxford, England: Blackwell.
- Macedonia, J. M., & Evans, C. S. (1993). Variation among mammalian alarm call systems and the problem of meaning in animal signals. *Ethology*, *73*, 177–197.
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge, England: Cambridge University.
- Marler, P. (1984). Animal communication: Affect or cognition? In K. R. Scherer & P. Ekman (Eds.), *Approaches to emotion* (pp. 345–365). Hillsdale, NJ: Erlbaum.
- Martin, P., & Bateson, P. (1986). *Measuring behavior: An introductory guide*. Cambridge, England: Cambridge University.
- McKinley, J., & Sambrook, T. D. (2000). Use of human-given cues by domestic dogs (*Canis familiaris*) and horses (*Equus caballus*). *Animal Cognition*, *3*, 13–22.
- Moelk, M. (1944). Vocalizing in the house-cat: A phonetic and functional study. *American Journal of Psychology*, *57*, 184–205.
- Owings, D. H., & Morton, E. S. (1998). *Animal vocal communication: A new approach*. Cambridge, England: Cambridge University.
- Owren, M. J., & Rendall, D. (1997). An affect-conditioning model of nonhuman primate vocal signaling. In D. H. Owings, M. D. Beecher, & N. K. Thompson (Eds.), *Perspectives in ethology: Vol. 12. Communication* (pp. 299–346). New York: Plenum.
- Rendall, D., & Owren, M. J. (2001). *No meaning required: Abandoning the search for a language grail in animal communication*. Manuscript submitted for publication.
- Rooney, N. J., Bradshaw, J. W. S., & Robinson, I. H. (2001). Do dogs

- respond to play signals given by humans? *Animal Behaviour*, 61, 715–722.
- Serpell, J. A. (2000). Domestication and history of the cat. In D. C. Turner & P. Bateson (Eds.), *The domestic cat: The biology of its behaviour* (pp. 179–192). Cambridge, England: Cambridge University.
- Shipley, C., Carterette, E. C., & Buchwald, J. S. (1991). The effects of articulation on the acoustical structure of feline vocalizations. *Journal of the Acoustical Society of America*, 89, 902–909.
- Smith, W. J. (1997). The behavior of communicating, after twenty years. In D. H. Owings, M. D. Beecher, & N. K. Thompson (Eds.), *Perspectives in ethology: Vol. 12. Communication* (pp. 7–53). New York: Plenum.
- Soproni, K., Miklósi, Á., Topál, J., & Csányi, V. (2001). Comprehension of human communicative signs in pet dogs. (*Canis familiaris*). *Journal of Comparative Psychology*, 115, 122–126.
- Struhsaker, T. T. (1967). Auditory communication among vervet monkeys (*Cercopithecus aethiops*). In S. A. Altmann (Ed.), *Social communication among primates* (pp. 332–354). Cambridge, England: Cambridge University.
- Tice, B., & Carrell, T. (2002). Sound format conversion software [Computer software]. Retrieved from <http://hush.unl.edu/LabResources.html>
- Tomasello, M., & Call, J. (1997). *Primate cognition*. New York: Oxford University.
- Vilá, C., Savolainen, P., Maldonado, J. E., Amorim, I. R., Rice, J. E., Honeycutt, R. L., et al. (1997, June 13). Multiple and ancient origins of the domestic dog. *Science*, 276, 1687–1689.
- Wilden, I., Herzel, H., Peters, G., & Tembrock, G. (1998). Subharmonics, biphonation, and deterministic chaos in mammalian vocalization. *Bioacoustics*, 9, 171–195.
- Zahavi, A. (1975). Mate selection: A selection for a handicap. *Journal of Theoretical Biology*, 53, 205–214.

Received August 22, 2001

Revision received March 4, 2002

Accepted March 6, 2002 ■

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 Demarie Jackson at 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 your 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 Demarie Jackson, Journals Office, American Psychological Association, 750 First Street, NE, Washington, DC 20002-4242.