Reliability and the adaptive utility of discrimination among alarm callers

Daniel T. Blumstein[*], Laure Verneyre and Janice C. Daniel

Department of Ecology and Evolutionary Biology, 621 Young Drive South, University of California, Los Angeles, CA 90095-1606, USA

Unlike individually distinctive contact calls, or calls that aid in the recognition of young by their parents, the function or functions of individually distinctive alarm calls is less obvious. We conducted three experiments to study the importance of caller reliability in explaining individual-discriminative abilities in the alarm calls of yellow-bellied marmots (*Marmota flaviventris*). In our first two experiments, we found that calls from less reliable individuals and calls from individuals calling from a greater simulated distance were more evocative than calls from reliable individuals or nearby callers. These results are consistent with the hypothesis that marmots assess the reliability of callers to help them decide how much time to allocate to independent vigilance. The third experiment demonstrated that the number of callers influenced responsiveness, probably because situations where more than a single caller calls, are those when there is certain to be a predator present. Taken together, the results from all three experiments demonstrate the importance of reliability in explaining individual discrimination abilities in yellow-bellied marmots. Marmots' assessment of reliability acts by influencing the time allocated to individual assessment and thus the time not allocated to other activities.

Keywords: alarm communication; individual recognition; reliability

1. INTRODUCTION

The alarm calls of some species have been reported to be individually distinctive (see, for example, Owings & Leger 1980; Leger *et al.* 1984; Cheney & Seyfarth 1988, 1990; Nikol'skii & Suchanova 1994; Blumstein & Armitage 1997; Hare 1998), but the function or functions of individually distinctive alarm calls is enigmatic (Blumstein 2005). Unlike the individually distinctive contact calls of primates (Rendall *et al.* 1996; Wanker & Fischer 2001) or the unique mother–offspring calls of species that leave their young in groups or colonies while foraging (Leonard *et al.* 1997; Jouventin *et al.* 1999; Insley 2000), alarm calls are vocalizations emitted under extreme duress (Klump & Shalter 1984) and might not initially be expected to have other social functions. Communication requires a receiver to respond differently in some way to signal variation (Bradbury & Vehrencamp 1998). Thus, we expect selection to act to allow signallers to create distinctive social calls, and we expect selection to allow receivers to discriminate between these social calls (e.g. Beecher 1982; Beecher *et al.* 1989). By contrast, it is not immediately obvious that selection acts to maximize the distinctiveness of alarm calls, but rather selection probably acts on receivers to discriminate among callers if by doing so receivers benefit. There are several, non-mutually exclusive functions for such discrimination.

First, as has been demonstrated with alarm calls in vervet monkeys (*Cercopithecus aethiops*; Cheney & Seyfarth 1988, 1990) and bonnet macaques (*Macaca radiata*; Ramakrishnan & Coss 2000), as well as in Richardson's ground squirrels (*Spermophilus richardsonii*; Hare & Atkins 2001),

California ground squirrels (*S. beecheyi*; Hanson & Coss 2001) and steppe marmots (*M. bobak*; Nesterova 1996), receivers could assess caller reliability. In these species, the alarm calls from unreliable individuals are less evocative than the calls from reliable individuals. Discrimination based on reliability might be particularly important in non-referential (Evans 1997) alarm calling systems where callers have variable thresholds to emit calls, and calls communicate the relative risk a caller was exposed to when calling (e.g. Nikolskii & Nesterova 1989, 1990; Harris *et al.* 1983; Blumstein & Armitage 1997; Blumstein 1999). Thus, a population of callers might include those that had a high false-alarm rate (i.e. low threshold to call), as well as those that had a low false-alarm rate (i.e. high threshold to call). In such situations, receivers should benefit by classifying individuals according to their reliability and modifying their responses accordingly.

Individual discrimination abilities, however, could be a by-product of selection for other functions (Blumstein 2005). For instance, if there are benefits from responding more to a certain age class or sex, or if it is important to determine whether more than one individual is calling, animals may in fact have the ability to discriminate among individuals.

Yellow-bellied marmots, a ground-dwelling social sciurid rodent, are an outstanding system to study the function of individual discrimination abilities. Yellow-bellied marmots emit two types of loud alarm call: the whistle and the trill (Blumstein & Armitage 1997). The whistle is the most commonly emitted alarm call type. Marmot whistles are not functionally referential, but rather they communicate the degree of risk a caller experiences (Blumstein & Armitage 1997; Blumstein 2003). Marmots emit more calls, louder calls, and call at a faster rate when risk is

higher. Importantly, acoustic characteristics of marmot whistles are individually distinctive (Blumstein & Armitage 1997). Discriminate function analysis also allows callers to be classified to sex or age at frequencies far above chance (Blumstein & Munos 2005). Playback experiments have demonstrated that marmots are able to discriminate among individuals based solely on their calls, and that calls from juveniles are more evocative than calls from other age-sex cohorts (Blumstein & Daniel 2005).

We aimed to study other possible functions of individual discrimination abilities in yellow-bellied marmots. First, using a training protocol where we experimentally created a 'reliable' and an 'unreliable' caller, we evaluated the hypothesis that individuals classify callers based on reliability. We found that marmots responded differently to reliable and unreliable callers, but in an unexpected way. Second, we artificially degraded calls to simulate callers calling from different distances and to create ambiguity about the real risk of predation. We assumed that a distant marmot's calls would be less reliable about the true risk of predation than calls from a nearby individual. We found that marmots responded differently to degraded and non-degraded calls. Finally, we directly tested the hypothesis that by discriminating among callers, marmots could determine whether one or more than one individual was calling. We thus demonstrated a novel function of individual discrimination abilities.

2. GENERAL METHODS

All experiments were conducted in and around the Rocky Mountain Biological Laboratory (RMBL), in Gothic, Colorado, USA, during the summers of 2002 and 2003. Marmots were live-trapped and individually marked with ear tags (for permanent identification) and non-toxic fur dye (for identification from afar; Armitage 1982).

For this study, we recorded alarm calls using Audix OM-3xb microphones (frequency response: between 40 Hz and 20 kHz) 20–40 cm from calling subjects, onto digital audio tape decks (Sony PCM-M1 or Tascam DA-P1) sampling at 44.1 kHz with 16 bit resolution from marmots contained in live traps. Our use of calls recorded on DAT equipment from trapped marmots ensured the highest quality of recorded calls. Moreover, because yellow-bellied marmots communicate risk, not predator type (Blumstein & Armitage 1997), we also controlled the context (and presumably the degree of risk) that calling marmots experienced. Alarm calls were acquired or transferred through a MOTU 828 Firewire external digital board (Mark of the Unicorn, Cambridge, MA, USA), to a Macintosh PowerBook G4 (Apple Computer, Cupertino, CA, USA), using CANARY 1.2 (Charif *et al.* 1995). Stimuli were then edited and normalized to match peak amplitudes in SOUNDEDIT 16 (Macromedia 1995), and transferred using s/pdif digital transfer protocols back to a Tascam DA-P1 for playback through Advent AV570 Powered Partners (Recoton Home Audio, Benicia, CA, USA; frequency response: between 40 Hz and 20 kHz). To our ears, the playbacks sounded natural and, as in previous experiments (e.g. Blumstein & Armitage 1997; Blumstein & Daniel 2005), marmots obviously responded to broadcast alarm calls by stopping behaviours they were engaged in, increasing vigilance, and rarely, by entering their burrows. For all experiments, we broadcast calls to non-pups (i.e. yearlings and adults) because we have no evidence from previous playback

studies that suggest systematic differences in how non-pups respond to playback.

For playback, we focused on the whistle—the most commonly emitted yellow-bellied marmot alarm call (Blumstein & Armitage 1997). Subjects were baited with a handful of Omolene 300 horse feed (Ralston Purina Inc., St Louis, MO, USA) to a location 1–2 m from their burrow. By baiting animals to a set location, and spreading out our playbacks throughout the morning active period, we attempted to target solitary marmots. We observed marmots from distances that did not obviously affect their behaviour (30–100 m, depending upon the individuals and the group). Alarm calls were broadcast from a speaker hidden 10–14 m from a burrow. Foraging marmots were video-recorded using a Cannon GL-1 mini-DV digital video recorder for one minute before beginning playback and during the playback itself. Because responses to playback may either be ephemeral or may be more appropriately measured over longer time-frames, we examined responses over several time-scales.

Videotapes were scored by an observer unfamiliar with the specific treatment using JWATCHER (Blumstein *et al.* 2000), where we noted the onset of each bout of foraging; standing quadrupedally and looking; rearing and looking while slouching bipedally on its hind legs; rearing up and looking while standing erect bipedally on its hind legs or toes; self-grooming; walking; running; and time spent in burrow. Focal records were analysed using JWATCHER, where we calculated the proportion of time allocated to foraging, heightened vigilance (rearing and rearing up and looking), normal vigilance (standing and looking), locomotion (walking and running) and time spent in the burrow. Because not all individuals engaged in all types of vigilance, we combined measures of heightened vigilance and normal vigilance into a total vigilance category. Ultimately, although we visually explored all the data, most formal analyses focused on the time allocated to foraging, because all subjects foraged before playback and responded to the playback by decreasing foraging. Although other responses were quantified, these responses were all interrelated. After playback, marmots mainly traded off foraging with vigilance, but they also spent some time either in locomotion or inside their burrows. Given that an animal that was not foraging could increase its vigilance or enter its burrow, for example, changes in vigilance *per se* were not as revealing as declines in foraging.

We calculated d, a measure of the effect size of pairwise comparisons, using the pooled standard deviation (Cohen 1988). For ANOVAs, we calculated the partial η^2 as a measure of effect size by using SPSS. By tradition, small effects have d-values of *ca.* 0.2, medium effects are *ca.* 0.5 and large effects exceed 0.8 (Cohen 1988).

(a) *Experiment 1: does caller reliability influence response?*

(i) *Methods*

The first experiment was designed to study the influence of caller reliability on responsiveness. We therefore conducted a learning experiment (*sensu* Hare & Atkins 2001) in which we paired the calls from one individual with the presence of a taxidermically mounted badger, a known predator of yellow-bellied marmots (Andersen & Johns 1977; Van Vuren & Armitage 1994), and the calls of another individual with a control stimulus (the badger covered with a tarp). Our control was chosen carefully. Both situations contained a three-dimensional stimulus, with presumably the same olfactory cues. Both had the same number of alarm calls played back during presentation, and both were set up and taken down identically. Thus, any response differences

can be attributed specifically to learning that the visual presence of the badger was paired with a particular individual's alarm calls. The individual whose calls were associated with the presence of the badger was defined as reliable because its calls were reliably associated with the presence of a predator. The calls of another individual that were emitted with no predator present were considered unreliable because these individuals called when there were no predators present.

This experiment potentially involved habituating animals to alarm calls by broadcasting them for 10 min throughout a marmot group. We thus viewed this experiment as potentially disturbing (Cuthill 1991) and wanted to minimize disturbance. To minimize our disturbance, we limited our sample size to 10 subjects and interpreted $0.1 < p < 0.05$ as potentially significant (Still 1982). Even with this relatively small sample size, we retained power by employing a within-subjects design and by standardizing the initial behaviour and the location of playback.

Specifically, we divided the learning experiment into three periods. The first period, typically lasting 2–3 days, was a pre-test period, during which we quantified the response of subjects to calls from the individuals who would later be reliable or unreliable. The second period, conducted on two consecutive days, was a training phase during which the four calls from one individual (the reliable one) were associated with the presence of the badger, and the four calls from the other individual (the unreliable one) with the presence of the covered badger. Groups were alternated with respect to whether they were exposed to the covered or uncovered badger first. We waited until as many animals from a group were out before conducting the training playbacks to ensure that as many individuals in the group were exposed to the reliable and unreliable calls. The third period, typically lasting 2–3 days, was a post-test period, during which we quantified the response of subjects to calls from the individuals who were either reliable or unreliable. For each of the periods, we used different exemplars from the same (unfamiliar) individuals.

For pre-test and post-test periods, the two playback stimuli included 1 min of silence followed with four different calls (four different exemplars were used during pre- and post-testing) emitted at a rate one call per second, for 4 s, followed by 1 min 56 s of silence. For the training playbacks, the two stimuli included four different exemplars of calls emitted at a rate one call per second for 10 min. Thus, we used 12 different exemplars, from four adult females, from four social groups (River South Mound, River Spruce Mound, Lower Picnic and Upper Picnic). Calls were played back to 10 non-pup subjects (three female yearlings, one male yearlings, five adult females, and one adult male) in the River Spruce Mound, River South Mound, Marmot Meadow Main Talus, Marmot Meadow Aspen Burrow, Bench and Gothic Townsite between 24 June and 1 August 2003. Animals responded to alarm calls by stopping foraging, engaging in vigilance and moving around. Although we analysed all responses, our planned statistical analyses focused primarily on the time allocated to foraging because all subjects foraged before playback and responded by decreasing foraging. Additionally, foraging and other behaviours are somewhat interrelated and, because an animal that suppressed foraging could increase its vigilance or enter its burrow, changes in vigilance, *per se*, are not as revealing as declines in foraging.

We compared, using a Wilcoxon matched-pairs signed rank test, the difference between post-test and pre-test for the reliable and unreliable caller. We focused both on the first 10 s of playback and the first minute after the start of the playback to compare both short-term, and longer-term responses. Although we aimed to
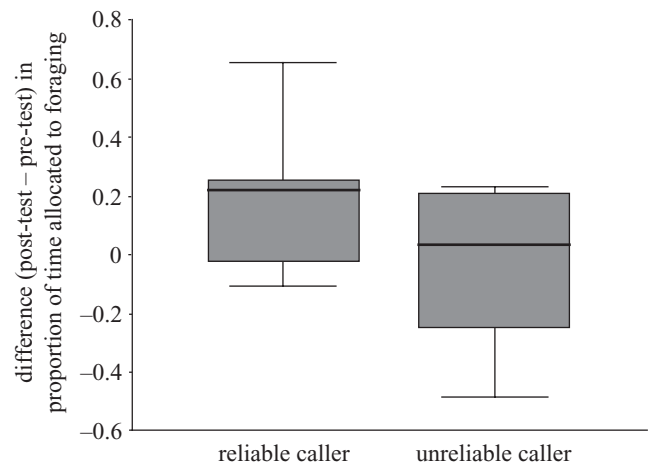
Figure 1. Box plots illustrating the difference (post-training – pre-training) in the proportion of time allocated to foraging in the 60 s during the playback of either a reliable individual or an unreliable individual. Reliability was manipulated directly by pairing, during a training period, the alarm calls of one individual with the presentation of a taxidermically mounted badger. The unreliable individual's alarm calls were not paired with the presence of a badger during the training period; $p = 0.066$.

have subjects hear only 'their' calls, more than a single subject was often within earshot. On average, an individual heard 3.23 (s.d. = 2.63, median = 3.5) playback series before each of their playbacks. We used linear regression to see if the number of playbacks heard could possibly confound our results. The average difference between the post-test and pre-test for the reliable callers was 5.57 days (s.d. = 1.50, median = 5.07 days), and 5.18 days (s.d. = 0.62, median = 5.01 days) for the unreliable callers.

(ii) *Results*

We found a moderately significant difference in the time allocated to foraging during the 1 min period following the onset of playback in the post-test period, compared with the pre-test period ($p = 0.066$, $n = 10$, $d = 0.850$). However, somewhat unexpectedly, marmots foraged more after hearing the reliable caller than the unreliable caller (figure 1). This difference was not found when analysing the first 10 s after playback ($p = 0.575$, $d = 0.425$). Total vigilance did not differ for the first 10 s after playback ($p = 0.878$) or for the entire 60 s after playback ($p = 0.333$).

There was no significant difference in the time allocated to foraging in the baseline period before playback ($p = 0.646$, $d = 0.094$), nor was there a significant effect of playback order on time allocated to foraging in the minute after playback ($p = 0.441$, $d = 0.263$). Our results were also not confounded by the number of other conspecifics at the bait ($R = 0.119$, $p = 0.466$), nor with the number of prior playbacks an individual was exposed to ($R = 0.006$, $p = 0.970$).

In conclusion, we found that during the first minute of playback, individuals demonstrated that they had learned about the reliability of the caller, and foraged more after hearing calls from reliable individuals. This suggests that reliability influences response, but in a novel way: after hearing calls from unreliable individuals, marmots had to independently assess predation risk.

(b) *Experiment 2: are distant calls more evocative than nearby calls?*

(i) *Methods*

If reliability influenced responsiveness by influencing the time an animal would allocate to independent investigation, then calls from adjacent marmots should be more reliable than those from distant individuals, because they presumably signify a predator that is threatening an adjacent caller. The aim of this experiment was to play back calls that were either acoustically degraded or non-degraded from different unfamiliar individuals (animals in different acoustically isolated social groups) to see if degraded calls were more evocative than non-degraded calls. Calls were degraded by broadcasting them through marmot habitat, and then by re-recording them 10 m from the speaker. We then normalized the calls to match peak amplitudes, and played them back at constant amplitude so that the only acoustic features that differed between them were those associated with degradation. It was important to normalize amplitude because previous experiments have demonstrated that amplitude alone influences responsiveness (Blumstein & Armitage 1997). However, perceived amplitude can vary considerably as a function of the relative direction that a signaller is facing (i.e. for a given source amplitude, calls from a signaller facing away from a receiver will be perceived as a lower amplitude than calls from a signaller facing a receiver). Following Richards (1981) and Naguib & Wiley (2001), we reasoned that to estimate distance, marmots should assess degradational changes in call structure and our experiment was designed to evaluate this.

Calls were recorded from four adult females in two social groups (Lower Picnic and Upper Picnic). We used five different exemplars per individual for this experiment. Calls were played back to 13 non-pup subjects (11 adult females, and 2 adult males) in Marmot Meadow Main Talus, Marmot Meadow Aspen Burrow, River South Mound, River Spruce Mound, Stonefield South and Stonefield Main groups, between 25 July and 5 August 2002. Animals responded to alarm calls by stopping foraging, engaging in vigilance, and moving around and occasionally entering their burrow. While we analysed all responses, our formal statistical analyses focused primarily on the time allocated to foraging because all subjects foraged before playback and responded by decreasing foraging. We used Wilcoxon matched-pairs signed rank test to test for differences between the response to degraded and non-degraded calls.

(ii) *Results*

Degraded calls were more evocative than non-degraded calls. Marmots suppressed foraging significantly more after hearing degraded calls in the first 10 s of playback ($p = 0.008$, $d = 1.233$) as well as during the entire minute of playback ($p = 0.023$; $d = 0.834$; figure 2) than after hearing non-degraded calls. After hearing degraded calls, marmots significantly increased the combined variable of total vigilance and time spent in the burrow in the first 10 s after playback ($p = 0.016$), and the minute after playback ($p = 0.050$).

There was no significant difference in the time allocated to foraging in the baseline period before playback ($p = 0.101$, $d = 0.675$). Nor was there a significant effect of playback order on time allocated to foraging either in the first 10 s after playback ($p = 0.859$, $d = 0.113$), or in the minute after playback ($p = 0.875$, $d = 0.010$). Our results were also not confounded by the number of other conspecifics at the bait (10 s: $R = 0.242$, $p = 0.253$; 60 s: $R = 0.337$, $p = 0.092$), nor with the number of
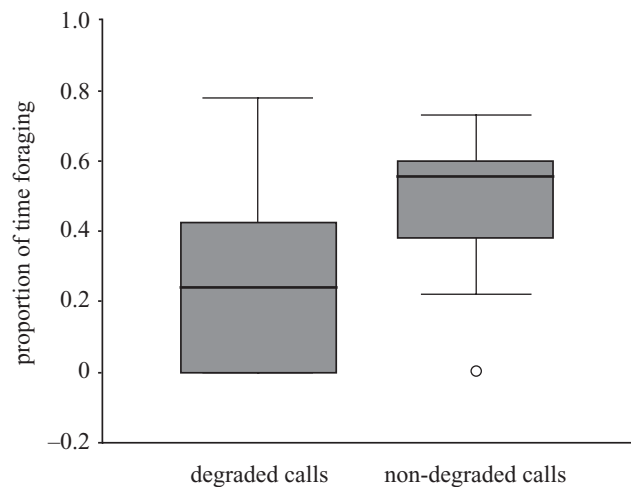
Figure 2. Box plots illustrating the proportion of time allocated to foraging during the 60 s playback period while hearing either non-degraded calls or calls degraded by broadcasting and then re-recording them 10 m from the speaker; $p = 0.023$.

prior playbacks an individual was exposed to (10 s: $R = 0.142$, $p = 0.489$; 60 s: $R = 0.252$, $p = 0.214$).

In conclusion, we found that individuals foraged less after hearing degraded calls. This suggests that reliability influences response, but again in a novel way: after hearing degraded alarm calls, which we assumed were less reliable because they communicated less certainty of risk to the perceiver, individuals allocated time to independently assess predation risk.

(c) *Experiment 3: can marmots identify multiple callers?*

(i) *Methods*

When predators move through a dense colony, more than a single marmot may call. An observational study of Belding's ground squirrels (*S. beldingi*) reported that multiple callers led to a greater colony level response (the number of individuals standing alert) than did single callers (Robinson 1981). The aim of this experiment was to play back calls from unfamiliar individuals (animals in acoustically isolated social groups) from different locations in a focal marmot's group to see if the calls from two individuals were more evocative than calls from a single individual. We selected eight whistles from adult females and used them to create three different stimuli. The first stimulus contained different exemplars from the same individual that were broadcast from only one speaker. Thus, the playback simulated a single non-moving caller. The second stimulus also contained different exemplars from one individual; however, the first four calls were emitted from one speaker, and the second four calls were emitted from a second speaker positioned 10–15 m away. Thus, the playback simulated a single individual calling from two distinct locations. The third stimulus contained calls from two individuals, each bout of calls emitted from a different speaker. Thus, the playback simulated two different callers in two different locations; a situation that should convey a relatively high risk. The three playback stimuli included 1 min of silence (baseline) followed by four calls emitted in 2 s, followed by four more calls 8 s later, followed by 1 min 50 s of silence.

Calls were played back to 20 non-pup subjects (nine female yearlings, five male yearlings, five adult females and one adult male) in the River South Mound, River Spruce Mound, Bench,

Marmot Meadow Main Talus, Marmot Meadow Aspen burrow and Gothic Townsite between 4 June and 8 July 2003. Each subject received all three stimuli in a counterbalanced, repeated-measures design. Ultimately, although we visually explored all the data, formal analyses focused on the time allocated to foraging because all subjects foraged before playback. For the analysis, we subtracted the time allocated to each behaviour in the first 10 s from the second 10 s after hearing the first stimulus and the second stimulus. We did this to see if animals habituated to the first caller. Positive values would imply habituation whereas negative or zero values would imply that marmots retained vigilance while hearing the second set of calls. We used a non-parametric Friedman ANOVA to test for overall response differences, followed by *post hoc* Wilcoxon matched-pairs signed rank tests. The key comparison was between the second and third stimulus because this tested directly whether the calls from two individuals were more evocative than from one individual.

We aimed to have subjects hear only their 'own' playback stimuli, but sometimes, more than a single individual was at the bait. We used linear regression to determine if this could possibly confound our results. On average, an individual heard 2.1 other playbacks (s.d. = 1.99, median = 2) before each of their playbacks, including the other playbacks directed to an individual. The average interval between the first and the second playback was 46.5 h (s.d. = 38.7 h, median = 24.5 h) and 123.5 h (s.d. = 165 h, median = 25 h) between the second and the third playback. We also used linear regression to determine if there was a relationship between our dependent variables of interest, and the number of other conspecifics.

(ii) Results

Marmots responded differently to the three playback treatments (Friedman non-parametric ANOVA, $p = 0.007$, $n = 20$, partial $\eta^2 = 0.197$; figure 3). *Post hoc* analysis revealed a significant difference in the proportion of time allocated to foraging between treatments 2 and 3 ($p = 0.001$, $d = 1.091$), but no significant difference between treatments 1 and 2 ($p = 0.126$, $d = 0.444$), or between treatments 1 and 3 ($p = 0.227$, $d = 0.334$). There was a significant increase in vigilance during playback. Specifically, after hearing two callers from two speakers, marmots allocated significantly more time to vigilance or time spent in their burrow ($p = 0.005$) compared with hearing the same individual from two speakers.

There was no significant difference in the time allocated to foraging in the baseline period before playback (Friedman $p = 0.549$, partial $\eta^2 = 0.045$). Nor was there a significant effect of playback order on time allocated to foraging (Friedman $p = 0.687$, partial $\eta^2 = 0.021$). Our results were not confounded by the number of conspecifics at the bait ($R = 0.073$, $p = 0.579$), nor with the number of prior experiments that an individual heard ($R = 0.123$, $p = 0.351$).

In summary, we have replicated the finding that marmots can differentiate individuals based solely on their calls (Blumstein & Daniel 2004), and we found that calls from two callers are more evocative than calls emitted from a single individual. This finding suggests that calls from two individuals denote a higher risk.

3. DISCUSSION

Marmots are able to discriminate callers based on reliability. After hearing calls from novel individuals that were made artificially unreliable, marmots foraged less in
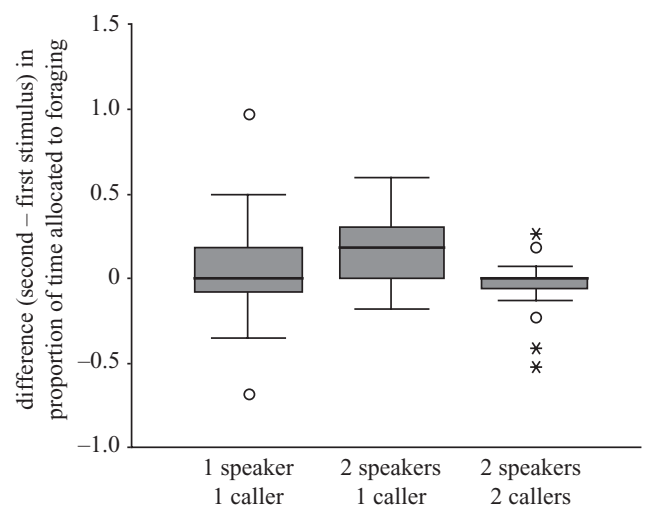
Figure 3. Box plots illustrating the difference (second stimulus – first stimulus) in time allocated to foraging after hearing alarm calls from either the same individual in one speaker, the same individual in two different speakers, or different individuals in two different speakers; $p = 0.007$.

the minute after playback than after hearing calls from individuals whose reliability was artificially created. This counterintuitive response is different from that reported in vervet monkeys (Cheney & Seyfarth 1988, 1990) and Richardson's ground squirrels (Hare & Atkins 2001), both of whom were less vigilant in response to unreliable callers. Our marmot results suggest that individuals may, after discriminating among callers, make their own independent assessment of relative risk. In addition, their previously formed 'concept of reliability' (Hare & Atkins 2001) influences the time they allocate to risk assessment. After hearing a reliable individual, and presumably not detecting a nearby predator, they resume foraging. Unreliable individuals, however, elicit further independent investigation.

The results of the second experiment are consistent with this general hypothesis about the importance of individual assessment after hearing an ambiguous call. After hearing degraded calls, individuals suppressed foraging longer and were more vigilant than after hearing non-degraded calls. Because calls were normalized for amplitude and their playback amplitudes matched it is, specifically, the difference in apparent distance communicated by relative degradation that was salient. Thus, marmots hearing a non-degraded call, presumably independently, verified that there was in fact no predator present, whereas marmots hearing a 'distant' degraded call allocated more time to independent assessment.

Thus, the results of the first two experiments demonstrate that reliability is important to marmots, but not in the way typically assumed. Variation in reliability influences the amount of time marmots individually assess predation risk. Reliable callers, or reliable situations, elicit less independent investigation than unreliable callers or unreliable situations.

The third experiment demonstrates a novel function of individual discrimination abilities: callers can determine whether more than one caller is calling. When predators pass through a large marmot colony, more than a single

individual may emit alarm calls. From the receiver's perspective, hearing calls from a single caller should be a less reliable cue to the real risk of predation than hearing calls from multiple callers. Our results demonstrate that calls from two callers are more evocative than calls from a single caller.

Taken together, the results from all three experiments demonstrate the importance of reliability in explaining individual discrimination abilities in yellow-bellied marmots. Marmots' assessment of reliability acts by influencing the time allocated to individual assessment and thus the time not allocated to other activities. This conclusion is also consistent with the previous observation that calls from juvenile marmots are more evocative than calls from adults (Blumstein & Daniel 2005).

However, and also consistent with playback results presented in Blumstein & Daniel (2005), which demonstrated no significant response differences to played back calls from different age–sex classes, we believe that reliability assessment was driven by the need to differentiate among individuals, rather than age or sex classes, for the following reasons. We tabulated the observations of alarm calls from 105 individuals (12 juvenile females, 10 juvenile males, 19 yearling females, 16 yearling males, 38 adult females, 10 adult males) who were identified calling during focal group observations (mean $\pm$ s.d. calls per individual: 3.2 ± 3.76) during five summers (1994, 1995, 2001, 2002, 2003). We classified a call as 'reliable' two different ways: if the caller was observed calling in response to an identified predator, or if the caller was observed calling in response to an identified predator or a human. We classified a call as 'unreliable' if the caller was observed calling in response to an obvious non-predator (e.g. deer; Blumstein & Armitage 1997). Many observations of calling were excluded because we were unable, with confidence, to identify the stimulus eliciting the call. For each individual, we calculated the proportion of calls that were reliable and averaged these values to generate a mean reliability per individual. We studied the effects on the proportion of reliable calls of age, sex and the interaction between age and sex using a two-way ANOVA. In all cases, we found no effect of age ($p_{\text{predators only}} = 0.444$; $p_{\text{predators + humans}} = 0.391$), sex ($p_{\text{predators only}} = 0.612$; $p_{\text{predators + humans}} = 0.157$), or the interaction of age × sex ($p_{\text{predators only}} = 0.536$; $p_{\text{predators + humans}} = 0.307$) on caller reliability. Thus, in yellow-bellied marmots, individuals, not age or sex classes, differ in their reliability. It follows that any selection for reliability assessment must be based on assessing the reliability of individuals and not age or sex classes.

Our results suggest that the assessment of reliability is a generally important reason why animals respond differently to individually distinctive alarm calls (Cheney & Seyfarth 1988, 1990; Hare & Atkins 2001) or to calls from less reliable classes of individuals (e.g. juveniles; Gouzoules *et al.* 1996; Nesterova 1996; Ramakrishnan & Coss 2000; Hanson & Coss 2001; Blumstein & Daniel 2005). At this point, we do not understand why reliability assessment may select for diametrically opposed responses, but we should generally expect considerable evolutionary flexibility in mechanisms of communication and assessment (Gerhardt & Huber 2002). Although these results strongly suggest that selection has acted on receivers to assess the reliability

of individuals producing alarm calls, it remains to be demonstrated whether selection has similarly acted on signallers to have distinctive vocalizations (Beecher *et al.* 1989), or whether individually distinctive vocalizations are a simple non-selected by-product of variation in vocal tract morphology (e.g. Fitch & Hauser 2003).

REFERENCES

Andersen, D. C. & Johns, D. W. 1977 Predation by badger on yellow-bellied marmot in Colorado. *Southwest. Nat.* **22**, 283–284.

Armitage, K. B. 1982 Yellow-bellied marmot. In *CRC handbook of census methods for terrestrial vertebrates* (ed. D. E. Davis), pp. 148–149. Boca Raton, FL: CRC Press.

Beecher, M. D. 1982 Signature systems and kin recognition. *Am. Zool.* **22**, 477–490.

Beecher, M. D., Loesche, P., Stoddard, P. K. & Medvin, M. B. 1989 Individual recognition by voice in swallows: signal or perceptual adaptation? In *The comparative psychology of audition: perceiving complex sounds* (ed. R. J. Dooling & S. H. Hulse), pp. 277–292. Hillsdale, NJ: Erlbaum.

Blumstein, D. T. 1999 Alarm calling in three species of marmots. *Behaviour* **136**, 731–757.

Blumstein, D. T. 2003 Social complexity but not the acoustic environment is responsible for the evolution of complex alarm communication. In *Adaptive strategies and diversity in marmots* (ed. R. Ramousse, D. Allaine & M. Le Berre), pp. 31–38. Lyon: International Network on Marmots.

Blumstein, D. T. 2005 The evolution of alarm communication in rodents: structure, function, and the puzzle of apparently altruistic calling. In *Rodent societies* (ed. J. O. Wolff & P. W. Sherman) (In the press.). University of Chicago Press.

Blumstein, D. T. & Armitage, K. B. 1997 Alarm calling in yellow-bellied marmots: I. The meaning of situationally-specific calls. *Anim. Behav.* **53**, 143–171.

Blumstein, D. T. & Daniel, J. C. 2005 Yellow-bellied marmots discriminate between the alarm calls of individuals and are more responsive to the calls from juveniles. *Anim. Behav.* **69**. (In the press.)

Blumstein, D. T. & Munos, O. 2005 Individual and age/sex class variation in marmot alarm calls. *Anim. Behav.* (In the press.)

Blumstein, D. T., Evans, C. S. & Daniel, J. C. 2000 JWATCHER 0.9. An introductory user's guide. http://www.jwatcher.ucla.edu.

Bradbury, J. W. & Vehrencamp, S. L. 1998 *Principles of animal communication*. Sunderland, MA: Sinauer.

Charif, R. A., Mitchell, S. & Clark, C. W. 1995 CANARY 1.2 user's manual. Ithaca, NY: Cornell Laboratory of Ornithology.

Cheney, D. L. & Seyfarth, R. M. 1988 Assessment of meaning and the detection of unreliable signals by vervet monkeys. *Anim. Behav.* **36**, 477–486.

Cheney, D. L. & Seyfarth, R. M. 1990 *How monkeys see the world*. University of Chicago Press.

Cohen, J. 1988 *Statistical power analysis for the behavioral sciences*, 2nd edn. Hillsdale, NJ: Erlbaum.

Cuthill, I. 1991 Field experiments in animal behaviour: methods and ethics. *Anim. Behav.* **42**, 1007–1014.

Evans, C. S. 1997 Referential communication. *Persp. Ethol.* **12**, 99–143.

Fitch, W. T. & Hauser, M. D. 2003 Unpacking 'honesty': vertebrate vocal production and the evolution of acoustic signals. In *Springer handbook of auditory research* (ed. A. Simmons, R. R. Fay & A. N. Popper), pp. 65–137. New York: Springer.

Gerhardt, H. C. & Huber, F. 2002 *Acoustic communication in insects and anurans*. University of Chicago Press.

Gouzoules, H., Gouzoules, S. & Miller, K. 1996 Skeptical responding in rhesus monkeys (*Macaca mulatta*). *Int. J. Primatol.* **17**, 549–568.

Hanson, M. T. & Coss, R. G. 2001 Age differences in the response of California ground squirrels (*Spermophilus beecheyi*) to conspecific alarm calls. *Ethology* **107**, 259–275.

Hare, J. F. 1998 Juvenile Richardson's ground squirrels, *Spermophilus richardsonii*, discriminate among individual alarm callers. *Anim. Behav.* **55**, 451–460.

Hare, J. F. & Atkins, B. A. 2001 The squirrel that cried wolf: reliability detection by juvenile Richardson's ground squirrels (*Spermophilus richardsonii*). *Behav. Ecol. Sociobiol.* **51**, 108–112.

Harris, M. A., Murie, J. O. & Duncan, J. A. 1983 Responses of Columbian ground squirrels to playback of recorded calls. *Z. Tierpsychol.* **63**, 318–330.

Insley, S. J. 2000 Long-term vocal recognition the northern fur seal. *Nature* **406**, 404–405.

Jouventin, P., Aubin, T. & Lengagne, T. 1999 Finding a parent in a king penguin colony: the acoustic system of individual recognition. *Anim. Behav.* **57**, 1175–1183.

Klump, G. M. & Shalter, M. D. 1984 Acoustic behaviour of birds and mammals in the predator context. I. Factors affecting the structure of alarm signals. II. The functional significance and evolution of alarm signals. *Z. Tierpsychol.* **66**, 189–226.

Leger, D. W., Berney-Key, S. D. & Sherman, P. W. 1984 Vocalizations of Belding's ground squirrels (*Spermophilus beldingi*). *Anim. Behav.* **32**, 753–764.

Leonard, M. L., Horn, A. G., Brown, C. R. & Fernandez, N. J. 1997 Parent–offspring recognition in tree swallows, *Tachycineta bicolor*. *Anim. Behav.* **54**, 1107–1116.

Macromedia. 1995 *SOUNDEDIT 16*. San Francisco, CA: Macromedia, Inc.

Naguib, M. & Wiley, R. H. 2001 Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. *Anim. Behav.* **62**, 825–837.

Nesterova, N. L. 1996 Age-dependent alarm behaviour and response to alarm call in bobac marmots (*Marmota bobac* Müll.). In *Biodiversity in marmots* (ed. M. Le Berre, R. Ramousse & L. Le Guelte), pp. 181–186. Moscow–Lyon: International Network on Marmots.

Nikol'skii, A. A. & Suchanova, M. V. 1994 Individual variability of alarm call in steppe marmot (*Marmota bobac* Müll., 1776). In *Actual problems of marmots investigation* (ed. V. Y. Rumiantsev), pp. 169–181. Moscow: ABF Publ. House.

Nikolskii, A. A. & Nesterova, N. L. 1989 Responses of *Marmota baibacina* to con- and heterospecific vocalizations. *Vest. Moscow Univ.* **16**, 50–55.

Nikolskii, A. A. & Nesterova, N. L. 1990 Perception of acoustic signals by *Marmota caudata* (Sciuridae, Rodentia). *Vest. Moscow Univ.* **16**, 53–59.

Owings, D. H. & Leger, D. W. 1980 Chatter vocalizations of California ground squirrels: predator- and social-role specificity. *Z. Tierpsychol.* **54**, 163–184.

Ramakrishnan, U. & Coss, R. G. 2000 Age differences in the responses to adult and juvenile alarm calls by bonnet macaques (*Macaca radiata*). *Ethology* **106**, 131–144.

Rendall, D., Rodman, P. S. & Emond, R. E. 1996 Vocal recognition of individuals and kin in free-ranging rhesus monkeys. *Anim. Behav.* **51**, 1007–1015.

Richards, D. G. 1981 Estimation of distance of singing conspecifics by the Carolina wren. *Auk* **98**, 127–133.

Robinson, S. R. 1981 Alarm communication in Belding's ground squirrels. *Z. Tierpsychol.* **56**, 150–168.

Still, A. W. 1982 On the number of subjects used in animal behaviour experiments. *Anim. Behav.* **30**, 873–880.

Van Vuren, D. & Armitage, K. B. 1994 Survival of dispersing and philopatric yellow-bellied marmots: what is the cost of dispersal? *Oikos* **69**, 179–181.

Wanker, R. & Fischer, J. 2001 Intra- and interindividual variation in the contact calls of spectacled parrotlets (*Forpus conspicillatus*). *Behaviour* **138**, 709–726.