SI Appendix 1

Method

Subjects. Subjects were undergraduates at the University of California, Santa Barbara, with normal or corrected-to-normal vision. Exp 1: *n*=30; Exp 2: *n*=38; Exp 3: *n*=28; Exp 4: *n*=28; Exp 5: *n*=38.

Procedure. In each trial, a black fixation cross appeared in the middle of a 15-inch computer monitor for 500 ms. A scene was then presented for 250 ms followed by a white screen for 250 ms. The alternate version of the scene was then presented for 250 ms and again followed by a white screen for 250 ms (Fig. 1). This series of presentations was repeated until the subject indicated (by mouse click) whether there was a changing object in the scene. The response and its latency were both recorded by the computer. This process continued until subjects had viewed and responded to all 70 scenes. The scene order was randomly assigned for each subject. One-third of trials were catch trials, in which nothing in the scene changed; which photos were catch trials was randomized across subjects.

An independent set of 26 subjects saw the same scenes as subjects in Exps 1 and 2, with the target item circled. They rated how interesting each target object was, and how consistent it was with its surrounding scene, using 7-point scales (1 = not interesting, not consistent, 7 = highly interesting, highly consistent).

Stimuli. *Exps 1-4*. Seventy scenes were taken from a commercially available CD-ROM collection of digital images (for the full set, see SI Appendices 3-7). The target object in each scene was from one of the five categories. Scenes were complex and natural, so most contained items from nontarget categories as well (e.g., a scene with a person target might

include plants, animals, and both kinds of artifacts). Each category was represented by 14 scenes with a target object from that category. However, one item from the animal set was later discovered to have a confounding visual change and was excluded from consideration in all statistical analyses. The scenes included urban and rural settings for all the categories. The target objects were as follows. *People*: both sexes and various ages, in a variety of orientations with respect to the observer. *Animals:* mammals, reptiles, birds, and insects. *Plants:* mostly trees and shrubs, but some potted flowers, fruits, and vegetables. *Moveable/manipulable artifacts:* common human-made tools and vehicles, e.g., stapler, wheelbarrow, boat, car. *Fixed artifacts*: artifacts of fixed location, often large enough to be construed as topographical landmarks, e.g., building, windmill, flag.

Targets were rated as semantically consistent: the mean consistency rating was above the midpoint for each category, and ranged from 4.09 (plants) to 5.23 (moveable artifacts). Although plants were judged consistent, their mean rating was lower than for the other four categories; however, given that inconsistency recruits attention (1), this would bias the stimuli against finding an animate advantage.

The images were 27 cm in height, 20.2 cm in width, and viewed from a distance of approximately 50 cm. When the target object was removed from a scene, it was replaced with surrounding background. The target objects occurred in a diverse range of positions. The use of natural scenes constrained the majority of the target objects, regardless of category, to the lower half of the image. Targets were, on average, 2.2 cm wide by 2.6 cm high. The target objects ranged in size from 0.5 cm wide by 0.6 cm high (a person) to 6.2 cm wide by 7.4 cm high (a tree). There were no significant differences between the animate and inanimate stimuli with respect to the target objects' luminance ($P = 0.34$), size ($P = 0.08$), or eccentricity ($P = 0.92$).

Exp 4. A Gaussian blur function was applied to each scene from Exp 1, using Photoshop 5.5 at a radius setting of 6.0 pixels. Examples are shown in SI Appendix 2.

Exp 5. Ninety-six images were employed, 49 of which were drawn from the previous stimuli set and the remainder drawn from the same CD-ROM collection (for full set, see SI Appendices 8-11). The images were the same size and presented under the same viewing conditions as in the prior four experiments. The targets averaged 1.97 cm in width and 2.00 cm in height. The targets ranged in size from 0.52 cm wide and 0.37 cm high (a horse) to 2.22 cm wide and 7.44 cm high (a person). Again, there were no significant differences between the animate and inanimate stimuli in size ($P = 0.28$) or eccentricity ($P = 0.46$). Inanimate objects were significantly higher with respect to luminance ($P = 0.001$); this, however, would bias the stimuli against the animate monitoring hypothesis [all else equal, higher luminance evokes greater visual attention (2)].

Analyses

There is no change to detect in the first 500 ms (because the scene with a change has not yet appeared on the screen). Reported reaction times do not reflect that 500-ms period.

Preliminary analyses showed no difference by category in detection of deletion-addition and left-right orientation changes, so these two types of change trials were collapsed for further analyses. When comparing responses to different semantic categories, each subject served as his or her own control (paired *t* tests). Reported *P* values are two-tailed.

1. False alarm rates: Exp 1: 0.92% (19/2,070); Exp 2: 1.6% (41/2,622); Exp 3: 1.04% (20/1,932); Exp 4: 3.99% (77/1,932); Exp 5: 2.6% (96/3,648).

2. Animacy or interestingness of target? For this analysis, the dependent variable was the mean reaction time for each scene (collapsed over Exps 1 and 2). A stepwise multiple

regression shows that animacy accounts for 31.0% of variance in RT; adding interest ratings increases it only slightly, to 32.7% (*P* for ΔF = 0.20). In contrast, adding animacy at step 2 increases the variance explained significantly, from 19.2% (for interest only) to 32.7% (*P* for ΔF = 0.001). The same pattern holds for hit rates: animacy explains 22.5% of variance and adding interest ratings increases this nonsignificantly to 22.8% (*P* for ΔF = 0.60). In contrast, adding animacy at step 2 increases variance explained from 10.8% to 22.8% (*P* for ΔF = 0.002).

3. Inversion (Exp 3): RT *M* = 5985 (SD 2,043), Exp 1 and 2 vs. Exp 3: *P* = 0.0003; accuracy *M* = 77.6% (SD 10.8), *P* = 0.64 (compared to upright, accuracy was worse for inverted animates but not for inverted inanimates). Low-pass filtered (Exp 4): RT *M* = 5,977 (SD 2,161), Exps 1 and 2 vs. Exp 4: *P* = 0.0008; accuracy *M* = 52.1% (SD 12.7), Exps 1 and 2 vs. Exp 4: $P = 10^{-17}$.

4. When inverted, moveable/manipulable artifacts were detected more slowly and less accurately than other inanimate objects. Despite this, there still was no overall animate RT advantage in Exp 3. If inverted moveable artifacts are considered anomalous and excluded from the analysis [yielding RT *M* = 5,246 (SD 1,925) for inanimates], the lack of an animate bias for inverted scenes is even more apparent: *P* = 0.65.

5. Because accuracy for inverted moveable/manipulable artifacts was disproportionately low, changes to inverted animate targets were detected more frequently than changes to inverted inanimate ones, taken as a group. But, as Fig 4*a* shows, this did not reflect a general animate advantage. It was caused by worse accuracy for inverted moveable artifacts compared to all other categories, including other inanimate targets [within inanimates: moveable vs. plants+fixed, $P = 10^{-6}$, *r* = 0.80]. If inverted moveable artifacts are considered anomalous and eliminated from analyses, the accuracy figures for inverted inanimate targets (plants and

fixed artifacts) and animate targets are about the same: animate $M = 77.7\%$ (SD 14.0), inanimate $M = 74.5\%$ (SD 17.6), $P = 0.21$.

6. There was no difference in the overall pattern of reaction times between the inversion and blur conditions (ANOVA, two conditions (inversion, blur) × five semantic categories: no main effect of condition: $P = 0.82$). As expected, however, their pattern was different from that for the upright, clear scenes, due to the animate advantage in the upright scenes [2 × 5, main effect of condition (Exps 1 and 2 vs. Exps 3 and 4): $P = 10^{-5}$, eta = 0.38)].

7. *Controlling for scene background.* For reasons of ecological validity, complex natural scenes are most appropriate for testing for an animate attentional advantage. This entails detecting a target in the context of a background scene. Differences in change detection as a function of whether the target's background scene is distracting or "busy" have not been reported in the literature. Nevertheless, we thought it would be prudent to test whether some unknown confound in scene backgrounds is driving the effects that we are attributing to target animacy.

Inversion shows that incidental differences in how "busy" the scene background is due to low level features cannot explain the animate attention advantage. But what about busyness due to high-level object recognition? One could imagine, for example, that changes to a target might be more difficult to detect when the background is cluttered with objects. Equally, detecting changes to a target may also be more difficult when the background contains interesting objects that compete for attention with the target. Indeed, if category-driven attentional effects exist, as we are claiming, then changes to a target might be more difficult to detect when there are animals or people in the background scene. This last possibility underlines an important point: How busy or interesting a scene is depends on properties of the observer's attentional system, many of which are still unknown. For this

reason, subjective ratings or measures that reflect the operation of the attentional system are needed to quantify how busy or interesting a background scene is.

To control for potential effects of this kind, 52 subjects were asked to rate scene backgrounds, that is, upright scenes with the targets absent (these were the "deletion" scenes used for the deletion-addition condition of the change detection experiment; thus surrounding background filled the space where the target had been). The subjects were drawn from the same population, but none had rated targets or participated in the change detection experiments. Twenty-six subjects rated the 70 scene backgrounds used in Exps 1 and 2; the other 26 rated the 96 scene backgrounds used in Exp 5. Using a 1-7 scale (1 = not at all), subjects rated each scene on "how busy" it is; after cycling through all the scenes they also rated each on "how interesting" it is (with busy-interesting order counterbalanced across subjects).

Most scene backgrounds were not viewed as very busy or interesting (mean ratings were at or below the scale midpoint, most ranging from 2.8-3.8). Regression analyses were conducted in which the dependent variable was either (*i*) the mean reaction time for detecting the target in a scene, or (*ii*) the mean hit rate for the target in a scene. (Because Exps 1 and 2 were identical, values for those scenes were computed from responses of all subjects in those experiments.) The three independent variables were (*i*) whether the target was animate or not, (*ii*) how busy the scene background was, and (*iii*) how interesting the scene background was. The goal was to determine whether reaction times were faster and hit rates higher for animate than inanimate targets, after controlling for how busy and how interesting the scene backgrounds were. Partial correlations show the unique effects of each variable, when all the others have been controlled for.

Results for Exps 1 and 2. After controlling for potential differences in scene background, animate targets still elicited significantly faster reaction times and higher hit

rates than inanimate targets [RT: partial r = -0.57, $P = 10^{-6}$ (sr = -0.55). Hits: partial r = 0.46, $P = 10^{-4}$ (sr = 0.45)]. β coefficients show that this corresponds to an advantage of 2,322 ms and 15.5 percentage points for animates, controlling for background. In contrast, there were no significant effects of scene background on the speed or accuracy with which targets were detected, either zero order or after controlling for animacy (RT, hits: for busy, Ps = 0.12, 0.22; for interesting, Ps = 0.80, 0.21).

Results for Exp 5. Background effects cannot account for the animate attentional advantage found in Exp 5 either. After controlling for differences in scene background, the advantage in speed and accuracy for animate over inanimate targets remained large and significant in Exp 5 [RT: partial r = -0.59, $P = 10^{-9}$ (sr = -0.53). Hits: partial r = 0.64, $P = 10^{-11}$ (sr = 0.59)]. Based on β coefficients, this corresponds to an advantage of 2,040 ms and 27 percentage points for animates over inanimates.

Non-human animals versus vehicles, Exp 5. The contrast between non-human animals and vehicles is important to our argument that the animate attentional advantage is produced by a phylogenetically ancient evolved mechanism, rather than by domain-general expertise. We therefore wanted to confirm that changes to non-human animals are detected faster and more accurately than changes to vehicles, after the potential effects of background busyness and background interestingness are statistically removed. (In the regression above, animate targets included people as well as non-human animals, and inanimates included artifacts in addition to vehicles.) To address this question, we conducted regression analyses in which the only animate targets were non-human animals and the only inanimate targets were vehicles. The results remained the same: Changes to non-human animals were detected faster and more accurately than changes to vehicles (with large effect size), even after controlling for differences in scene background [RT: partial r = -0.65, $P = 10^{-6}$ (sr = -0.54). Hits: partial r = 0.56, P = 0.00006 (sr = 0.53)]. The attentional advantage for non-human

animals over vehicles, controlling for scene background, corresponds to 1,492 ms and 24 percentage points. This shows that non-human animals are detected faster than vehicles, and that this difference cannot be explained by incidental differences in scene backgrounds.

Should future researchers monitor scene background? Future researchers designing change detection experiments with complex natural scenes may be interested in whether they need to take account of scene background in their experimental designs. Controlling for whether the target was animate, scene background had no independent effects on change detection for the scenes used in Exps 1 and 2, but it did for the scenes used in Exp 5. After controlling for all other variables in Exp 5, busyness of background was correlated with increased reaction time and decreased accuracy in detection of targets [RT: partial $r = 0.38$, $P = 0.00014$ ($sr = 0.30$). Hits: partial $r = -0.33$, $P = 0.001$ ($sr = -0.25$)]. Surprisingly, how interesting the scene background was exerted an effect in the opposite direction from busyness: Controlling for busyness and animacy, targets were not detected more accurately, but they were detected faster, when the scene background was more interesting [RT: partial $r = -0.32$, $P = 0.0017$ ($sr = -0.24$). Hits: partial $r = 0.17$, $P = 0.11$).

This means that how busy and how interesting a background scene is can affect the speed and accuracy with which changes to a target are detected, independent of that target's semantic category or other properties. Our analysis shows that background effects cannot explain the animate attentional advantage. But backgrounds should continue to be monitored in future research, because they can have an independent effect on change detection.

Conclusion, scene background analyses. The animate attentional advantage remains significant and large, even when controlling for how busy and how interesting the target's background scene is. This is true even when one compares non-human animals to vehicles.

8. *Controlling for low level visual properties in Exp 5*. As for Exps 1 and 2, we wanted to make sure that the animate detection advantage in Exp 5 was independent of any incidental differences in the low level visual properties of scenes.

Target size, eccentricity, and luminance were regressed onto the mean reaction time and hit rate for each scene in Exp 5. Target size and eccentricity did not predict scene reaction times or hit rates. Changes to less luminant targets were detected a little faster and more accurately in Exp 5 (RT: $P = 0.058$. Hits: $P = 0.031$). The literature consistently reports the opposite—that more luminant targets recruit attention (2), so the fact that change detection was slightly better for less luminant targets probably reflects the animate attentional advantage (animate targets were less luminant in Exp 5, see above).

To control for incidental differences due to all possible low level visual properties, we conducted a change detection experiment ($n = 31$) using inverted scenes from Exp 5 (analogous to Exp 3). Inversion disrupts high level object recognition while perfectly preserving all low level visual properties of the scenes.

That inversion disrupted target recognition is most evident from the decrease in hit rates compared to upright scenes of people (-32 points, from 92% upright to 60% inverted), animals (-24 points, 89% vs. 65%), and vehicles (-16 points, 63% vs. 47%). (Static artifacts: -7 points, from an (already low) figure of 59% vs. 52%.).

For the scenes used in Exps 1 and 2, inversion had eliminated the animate detection advantage. But the inverted scenes in Exp 5 yielded some animate-inanimate differences in reaction times and hit rates, though smaller than those found for the upright scenes of Exp 5. The hit rates for inverted people and non-human animals were comparable, but they were higher than those for inverted vehicles and static artifacts. Reaction times showed the same pattern: not different for inverted people and animals (3,578 and 3,377 ms, respectively), but

RTs for both animate categories were a little faster than those for inverted vehicles and static artifacts (3,989 and 3,983 ms, respectively).

Inversion disrupts high level object recognition, but does not wipe it out completely, so these differences for inverted scenes could represent the animate attentional advantage kicking in when an inverted person or animal is recognized as such. Alternatively, the advantage in change detection for inverted animals and people could represent nothing more than incidental differences in low level visual properties of the scenes in which they appeared. If so, then we must ask whether the animate attentional advantage found in Exp 5 is real, or is it merely an artifact of differences in low level visual features of the scenes we happened to use as stimuli?

To answer this question, we reanalyzed the data from Exp 5 (upright scenes) using the inversion results to control for low level visual features, and did so in a way that would maximally jeopardize the animate monitoring hypothesis. We did this by making the conservative assumption that all the differences in change detection between inverted scenes, including the differences between inverted animate and inverted inanimate targets—were due to differences in low level visual features of the scenes (and not to differences in animate monitoring). In this view, a scene's inversion score reveals the extent to which low level stimulus properties of that scene and target make it easier or more difficult to detect changes in the target. For the purposes of this analysis, the inverted target's semantic category (person, animal, vehicle, artifact) is assumed to play no role in change detection.

For each scene, the advantage or disadvantage in reaction time due to low level properties was quantified by calculating the extent to which the inverted scene's mean RT deviates from the mean RT for all inverted scenes (the grand mean). For example, a mean RT for inverted Scene A that is 150 ms slower than the mean RT for all inverted scenes would indicate a disadvantage in reaction time due to low level features. To correct for this

disadvantage, 150 ms would therefore be subtracted from each subject's RT for the upright Scene A they saw in Exp 5. Similarly, an inverted RT for Scene B that is 200 ms faster than the mean RT for all inverted scenes would indicate an advantage in reaction time due to low level features. To correct for this advantage, 200 ms would be added to each subject's RT for the upright Scene B in Exp 5. Applying these corrections to the results for the upright scenes in Exp 5 eliminates any advantage or disadvantage in change detection resulting from low level visual features.

The system for correcting hit rates was analogous, but modified to accommodate the fact that hits are binary (see below for details)*.

Note that this method of correcting for low level features is strongly biased against the animate monitoring hypothesis. It assumes that all differences in inverted scenes are due to low level features. In reality, however, it seems likely that some fraction of these differences result from animate attentional monitoring (given that at least some inverted targets will eventually be recognized as animals or people). Using inversion scores to correct for low level features therefore has the side-effect of also removing legitimate effects of animate monitoring in response to inverted targets from effects of animate monitoring in response to the upright targets in Exp 5.

Nevertheless, the animate attentional advantage remained large and significant even after the correction for low level features was applied to the results of Exp 5. Changes to animate targets were detected more than a second faster than changes to inanimate targets (2,717 ms vs. 3,978 ms, $r = 0.74$, $P = 10^{-7}$), and with much greater accuracy (hits: 88% vs. 63%, $r = 0.88$, $P = 10^{-12}$). Moreover, changes to non-human animals are still detected faster and more accurately than changes to vehicles, even when corrected scores are used (2,856 ms vs. 3,754 ms, $r = 0.42$, $P = 10^{-5}$. Hits 79% vs. 67%, $r = 0.56$, $P = 0.0002$).

The corrected scores by category were 2,578 ms and 97% hits for people; 2,856 ms and 79% hits for non-human animals; 3,754 ms and 67% hits for vehicles; 4,201 ms and 58% hits for static artifacts. The uncorrected scores for people and animals in Exp 5 were indistinguishable, but these corrected scores seem to indicate a detection advantage for people over non-human animals. A more likely interpretation, however, is that the visual system is designed such that animals in motion are particularly easy to recognize (and, therefore, likely to recruit attention), even in inverted scenes, which would bias the correction procedure disproportionately against non-human animals. Indeed, when scenes were inverted, changes to animals in motion were detected 400-800 ms faster and 11-25 percentage points more accurately than changes to inverted targets from any other category, including humans (inverted people in motion were next best, but still 400 ms slower and 11 points less accurate). Because the inversion correction removes real effects of animate monitoring along with nuisance effects of low level features, it will remove real effects of animate monitoring disproportionately from non-human animal targets precisely to the extent that inverted animals in motion are recognized and monitored better than other inverted targets.

**Details of low-level visual feature correction for hits*. Each subject either detects the change in a scene or not (a binary score 1 or 0 for upright scenes), so subtracting deviation scores for inverted scenes (e.g., +4 points, -7 points) would result in a measure without a direct interpretation as "percent of hits detected". So for hits, inversion results were used to calculate deviations at the category level, where a scene's category is defined by the target's semantic category [i.e., static person, dynamic person, static animal, dynamic animal, static artifact, dynamic artifact (i.e., vehicle)]. The correction factor was based on the extent to which the mean hit rate for a given category of inverted scenes deviates from the mean hit rate for all inverted scenes. For example, a mean hit rate for inverted "static artifact" scenes

that is 5 points *lower* than the mean hit rate for all inverted scenes would indicate a disadvantage in change detection for that category due to low level features. To correct for this disadvantage, 5 points would be added to each subject's mean hit rate for (upright) static artifacts in Exp 5. Likewise, a mean hit rate for inverted "static people" scenes that is 8 points higher than the mean hit rate for all inverted scenes would indicate an advantage in change detection for that category due to low level features. To correct for that low level advantage, 8 points would therefore be subtracted from each subject's mean hit rate for (upright) static people in Exp 5.

1. Hollingworth A, Henderson J (2000) *Visual Cognit* 7**:** 213-235.
2. Turatto M, Galfano G (2002) *Vision Res* 40: 1639-1644.