Category-specific attention for animals reflects ancestral priorities, not expertise

Joshua New*†‡, Leda Cosmides*, and John Tooby*

*Center for Evolutionary Psychology, University of California, Santa Barbara, CA 93106; and †Department of Psychology, Yale University, New Haven, CT 06520

Edited by Gordon H. Orians, University of Washington, Seattle, WA, and approved August 10, 2007 (received for review May 3, 2007)

Visual attention mechanisms are known to select information to process based on current goals, personal relevance, and lower-level features. Here we present evidence that human visual attention also includes a high-level category-specialized system that monitors animals in an ongoing manner. Exposed to alternations between complex natural scenes and duplicates with a single change (a change-detection paradigm), subjects are substantially faster and more accurate at detecting changes in animals relative to changes in all tested categories of inanimate objects, even vehicles, which they have been trained for years to monitor for sudden life-or-death changes in trajectory. This animate monitoring bias could not be accounted for by differences in lower-level visual characteristics, how interesting the target objects were, experience, or expertise, implicating mechanisms that evolved to direct attention differentially to objects by virtue of their membership in ancestrally important categories, regardless of their current utility.

animacy | category specificity | domain specificity | evolutionary psychology | visual attention

Visual attention is an umbrella term for the set of operations that select some portions of a scene, rather than others, for more extensive processing. These operations evolved because some categories of information in the visual environment were likely to be more important or time-sensitive than others for activities that contributed to an organism's survival or reproduction. The selection criteria that direct visual attention can be categorized by their origin: (*i*) goal-derived: criteria activated volitionally in response to a transient internally represented goal; (*ii*) ancestrally derived: criteria so generally useful for a species, generation after generation, that natural selection favored mechanisms that cause them to develop in a species-typical manner; and (*iii*) expertise-derived: criteria extracted during ontogeny by evolved mechanisms specialized for detecting which perceptual cues predict information that enhances task performance.

These three types of criteria may also interact; for example, differential experience or temporary goals could calibrate or elaborate ancestrally derived criteria built into the attentional architecture.

The ways in which human attention can be affected by goals and expertise have been extensively investigated. Indeed, humans are zoologically unique in the extent to which we evolved to engage in behavior tailored to achieve situation-specific goals as a regular part of our subsistence and sociality (1, 2). Among our foraging ancestors, improvising solutions in response to the distinctive features of situations would have benefited from the existence of goal-driven voluntary attentional mechanisms. As predicted by such a view, otherwise arbitrary but task-relevant objects command more attention than task-irrelevant ones (3), and expertise in a task domain shifts attention to more task-significant objects (4), features (5), and locations (6).

In contrast, attentional selection criteria that evolved in response to the payoffs inherent in the structure of the ancestral world have been less systematically explored. Yet, the rapid identification of the semantic category to which an object belongs (e.g., animal, plant, person, tool, terrain) and what its presence in the scene signifies [e.g., predatory danger, food (prey), offspring at risk] would have been central to solving many ancestral adaptive problems. That is, stably and cumulatively across hundreds of thousands of generations, attention allocated to different semantic categories would have returned different average informational payoffs. From this perspective, it would be odd to find that attention to objects was designed to be deployed in a category-neutral way. Yet there has been comparatively little research into whether some semantic categories spontaneously recruit more attention than others, and whether such recruitment might be based on evolved prioritization. Most exceptions have studied attention and responses to highly social information such as faces (7, 8), eye gaze (9), hand gestures (10), and stylized human outlines (stick drawings and silhouettes) (11).

The Animate Monitoring Hypothesis

For ancestral hunter-gatherers immersed in a rich biotic environment, non-human and human animals would have been the two most consequential time-sensitive categories to monitor on an ongoing basis (12). As family, friends, potential mates, and adversaries, humans afforded social opportunities and dangers. Information about non-human animals was also of critical importance to our foraging ancestors. Non-human animals were predators on humans; food when they strayed close enough to be worth pursuing; dangers when surprised or threatened by virtue of their venom, horns, claws, mass, strength, or propensity to charge; or sources of information about other animals or plants that were hidden or occluded; etc. Not only were animals (human and non-human) vital features of the visual environment, but they change their status far more frequently than plants, artifacts, or features of the terrain. Animals can change their minds, behavior, trajectory, or location in a fraction of a second, making their frequent reinspection as indispensable as their initial detection.

For these reasons, we hypothesized that the human attention system evolved to reliably develop certain category-specific selection criteria, including a set designed to differentially monitor animals and humans. These should cause stronger spontaneous recruitment of attention to humans and to non-human animals than to objects drawn from less time-sensitive or vital categories (e.g., plants, mountains, artifacts). We call this the animate monitoring hypothesis. Animate monitoring algorithms are hypothesized to have coevolved alongside goal-driven

Author contributions: J.N., L.C., and J.T. designed research; J.N. performed research; J.N. and L.C. analyzed data; and J.N., L.C., and J.T. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviations: CD, change detection; Exp *n*, Experiment *n*; RT, reaction time.

See Commentary on page 16396.

‡To whom correspondence should be addressed. E-mail: joshua.new@yale.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/0703913104/DC1.

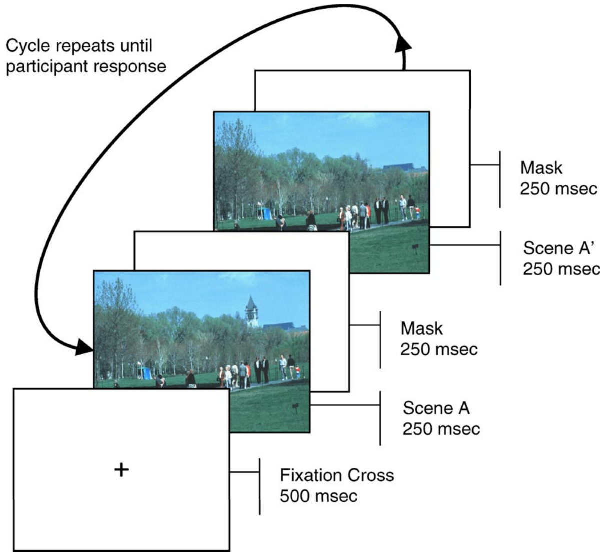


Fig. 1. Diagram illustrating the sequence and timing of each trial in Exp 1–5.

voluntary processes that focus attention on task-relevant objects, providing the voluntary system with one of several interrupt circuits made necessary by a surprising world. These algorithms should operate automatically and autonomously from executive function, so that important changes in non-humans and humans can be detected rapidly, even when they are unexpected or irrelevant to current goals or activities. Hence, we propose that animate inputs will recruit visual attention in a way that is less context-, goal-, expertise-, and state-dependent than other inputs. Although increasingly focused attention may increasingly screen out task-irrelevant stimuli, such exclusion should affect human and animal stimuli less than members of other categories. In particular, subjects' attention should display the predicted animate monitoring bias in the absence of instructions to look for animals or humans and regardless of their relevance to the task or to subjects' goals.

The counterhypothesis is that visual attention contains no mechanisms designed to differentially allocate attention on the basis of the semantic category of the input. This means there should be no mechanisms that evolved to deploy attention differentially to animate targets, and therefore no animate monitoring bias should be found. If, nevertheless, evidence of such a bias were to be found, the fallback hypothesis would be that such an effect would be the result of expertise: that is, starting with an equipotential attentional system, ontogenetic training would accrete attentional biases as a function of differential experience with the stimulus inputs and their ontogenetic importance. We will call this the expertise hypothesis.

Assessing Preferential Attention

Experiments show that viewers often fail to detect sizeable changes in an image when these occur during very brief interruptions of sight, a phenomenon known as change blindness (13, 14). To explore the selection criteria implemented by attentional mechanisms, we used the change detection (CD) paradigm (Fig. 1), in which viewers are asked to spot the difference between two rapidly alternating scenes that are identical except for a change to one object. The logic is straightforward: in a CD paradigm, changes to more attended objects or regions in a complex natural scene will be detected faster and more reliably than changes to less-attended ones. By varying which features in a scene are changed, one can learn the criteria by which visual attention mechanisms select objects for further processing. In a CD experiment, subjects are instructed to detect changes, but they are not given any task-specific goal that would direct their attention to some kinds of objects over others. Thus, the CD paradigm can be used to investigate how attention is deployed in the absence of a voluntary goal-directed search (15). If the animate bias hypothesis is correct, then change blindness will be attenuated for animals and humans compared with other object categories. This is because category-specific attention mechanisms will automatically check the status of animals and people on an ongoing basis.

We adapted a standard CD task (14) to test for the predicted category-specific biases (Fig. 1). The stimuli were color photographs of natural complex scenes (Fig. 2). For Experiments (Exp) 1–4, 70 scenes with target objects from five semantic categories were used (14 in each category): two animate (people and animals) and three inanimate [plants; moveable/manipulable artifacts designed for interaction with human hands/body (e.g., stapler, wheelbarrow); fixed artifacts construable as topographical landmarks (e.g., windmill, house)]. These categories were chosen because converging evidence from neuropsychology and cognitive development suggests each is associated with a functionally distinct neural substrate (16, 17). Each involves an evolutionarily important category, but only the animates require close visual monitoring. Target categories for Exp 5 (96 scenes) were vehicles, artifacts that do not move on their own, non-human animals, and people. [For details, see supporting information (SI) *Appendix 1*].

Tests and Predictions

If, as hypothesized, the human attentional architecture includes evolved mechanisms designed to differentially direct attention to both human and non-human animals, then, in a CD task using complex natural scenes, we predict that: (*i*) changes to animals (both human and non-human) will be detected more quickly than changes to inanimate objects and (*ii*) changes to animals will be detected more frequently than changes to inanimate objects. By hypothesis, attention is differentially recruited to animals by virtue of neural processes recognizing (at some level) their category membership. The bias is category-driven. Therefore, (*iii*) although animals will be judged more interesting than inanimate objects, detection rates will be better predicted by the target's category (animate or inanimate) than by how interesting the targets are judged to be, and (*iv*) the detection advantage for animate categories will not be due to lower-level perceptual characteristics, such as visual complexity or high contrast.

According to the expertise counterhypothesis, any effects by category will arise from differences in frequency of observation, differential training with different categories, or the relative importance during ontogeny of paying differential attention by category. We selected vehicles as an evolutionarily novel contrast category with which subjects have a great deal of experience; which move and do so in a self-propelled fashion; and which subjects have been trained from childhood as pedestrians and drivers to differentially attend to because of the life-or-death importance of anticipating their momentary shifts in trajectory. In comparison, our subjects see and interact with non-human animals far less often than with vehicles, and animals have little practical significance to our subjects. Despite greater subject expertise with vehicles, we predict that (*v*) the animate bias will not be a consequence of ontogenetic exposure to things in motion. In particular, although subjects see large numbers of vehicles moving every day, changes to vehicles will not be detected as quickly or reliably as changes to animals and people.

Finally, this study affords an opportunity to measure the effects of expertise on visual attention. Subjects have a lifetime of intensive training in attending to one species above all: humans. In contrast, subjects have orders-of-magnitude less experience attending to any other given species. The difference

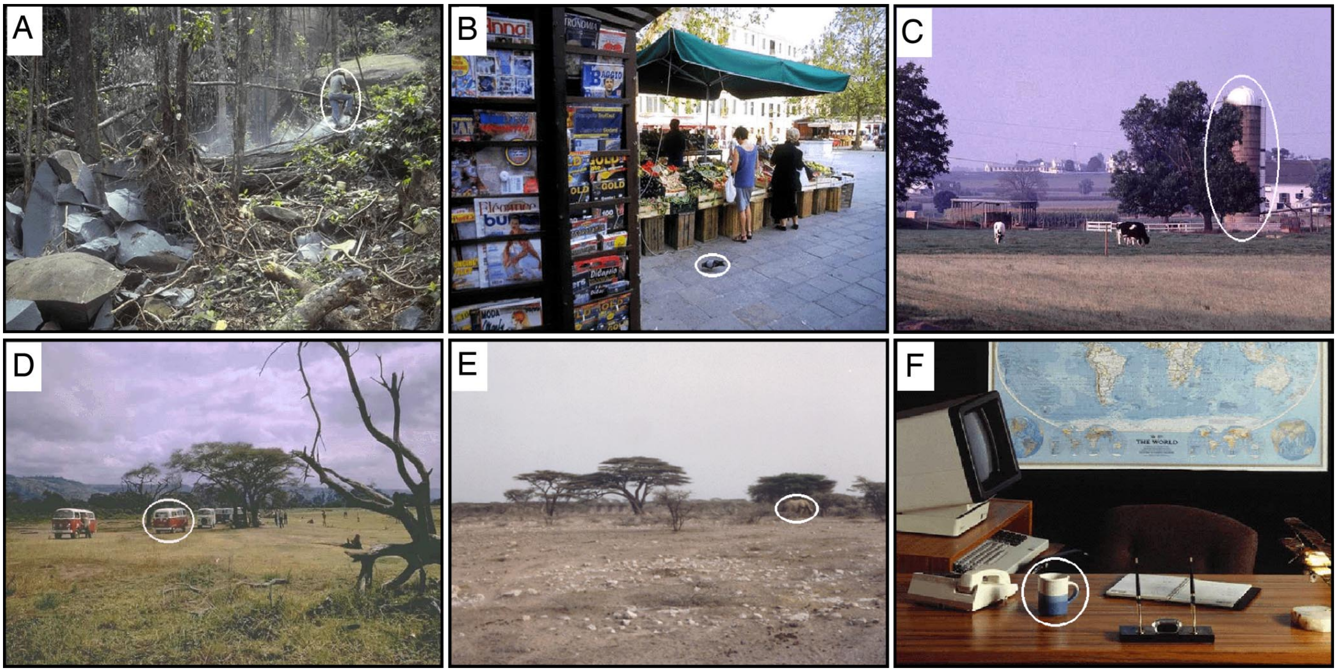


Fig. 2. Sample stimuli with targets circled. Although they are small (measured in pixels), peripheral, and blend into the background, the human (*A*) and elephant (*E*) were detected 100% of the time, and the hit rate for the tiny pigeon (*B*) was 91%. In contrast, average hit rates were 76% for the silo (*C*) and 67% for the high-contrast mug in the foreground (*F*), yet both are substantially larger in pixels than the elephant and pigeon. The simple comparison between the elephant and the minivan (*D*) is equally instructive. They occur in a similar visual background, yet changes to the high-contrast red minivan were detected only 72% of the time (compared with the smaller low-contrast elephant's 100% detection rate).

in performance between attention to humans and attention to other animal species gives a measure of the importance of expertise in training attention to animate inputs.

Results

Exp 1 was designed to test predictions *i–iii* (above) of the animate bias hypothesis, Exp 2 was a replication of Exp 1, and Exp 3–5 were designed to test predictions *iv–v*.

The hit rate (percent correct) was used to assess accuracy, because false alarms were so rare across the five experiments (2% of all responses; *SI Appendix 1.1*). Reaction times (RTs) are for hits.

Do Animals and People Recruit Preferential Attention? Yes. Changes to animals and people were detected more often and more quickly than changes to inanimate objects in Exp 1 and 2 (Fig. 3 *A* and *B*). More specifically, changes to animate targets (animals and people) were detected faster than changes to inanimate ones (plants, moveable artifacts, and fixed artifacts), both in Exp 1 and its replication (Exp 2); animate vs. inanimate target RTs: $P = 10^{-10}$ and 10^{-15}, respectively. Changes to animate targets were detected 1–2 seconds faster than changes to inanimate ones, and the effect size (r) associated with this difference was large in both experiments (0.88 and 0.86).

The greater speed in detecting changes to animals and people was not achieved at the expense of accuracy. On the contrary, subjects were faster and more accurate for animate targets, which elicited hit rates 21–25% points higher than inanimate targets (Exp 1 and 2, $r = 0.84$ and 0.80; $P = 10^{-8}$ and 10^{-10}; false-alarm rates were low, 0.92% and 1.6%). Overall, 89% of changes to animate targets were correctly detected vs. 66% of changes to inanimate ones. The animate advantage in speed and accuracy remains strong, even when inanimates are compared only to non-human animals (see Fig. 3; RT, $r = 0.80$ and 0.64, $P = 10^{-7}$ and 10^{-11}; hits, $r = 0.82$ and 0.63, $P \leq 0.0002$).

Following convention, we reported RTs for hits only. However, this measure fails to capture cases in which a change to the target was missed entirely; missing a change is a more severe case of "change blindness" than being slow to notice one. Subjects were change-blind more often for inanimate targets than for animate ones (miss rates, 34% inanimate vs. 11% animate). Because this is not reflected in mean RTs for hits, the difference between animate and inanimate RTs underestimates the animate attentional advantage. Moreover, mean RTs can mask important differences in the time course of change detection.

Fig. 3 addresses these concerns by showing, for each category, the time course of change detection. The relationship between time elapsed and total number of changes detected is plotted. Steeper slopes indicate earlier change detection; higher asymptotes mean more changes were eventually detected (i.e., less change blindness). Consistent with the hypothesis that animals and people should undergo incidental monitoring so that changes in their location and state can be rapidly detected, the curves for the two animate categories have steeper slopes and asymptote at higher levels than those for the three inanimate categories. Moreover, there appear to be attentional capture as well as monitoring effects.

Attentional Capture. The animate and inanimate curves diverge quickly: there were more hits for animate than for inanimate targets even for the fastest responses, ones in which changes were detected in <1 second (Exp 1, hits 8.8% vs. 3.9%; $P = 0.0025$, $r = 0.52$, no false alarms; Exp 2, hits, 3.8% vs. 1.6%, $P = 0.002$, $r = 0.48$; one false alarm). This suggests that animates capture attention in addition to eliciting more frequent monitoring. The maximal difference between animate and inanimate detection occurred at 3.5–4 elapsed seconds, a 33–37% point difference, with an effect size of $r > 0.93$ (P values $=10^{-14}$).

Do Animals and People Receive Preferential Attention Because They Are More "Interesting?" In CD studies, interesting items are detected faster than uninteresting ones (14, 18). When a separate

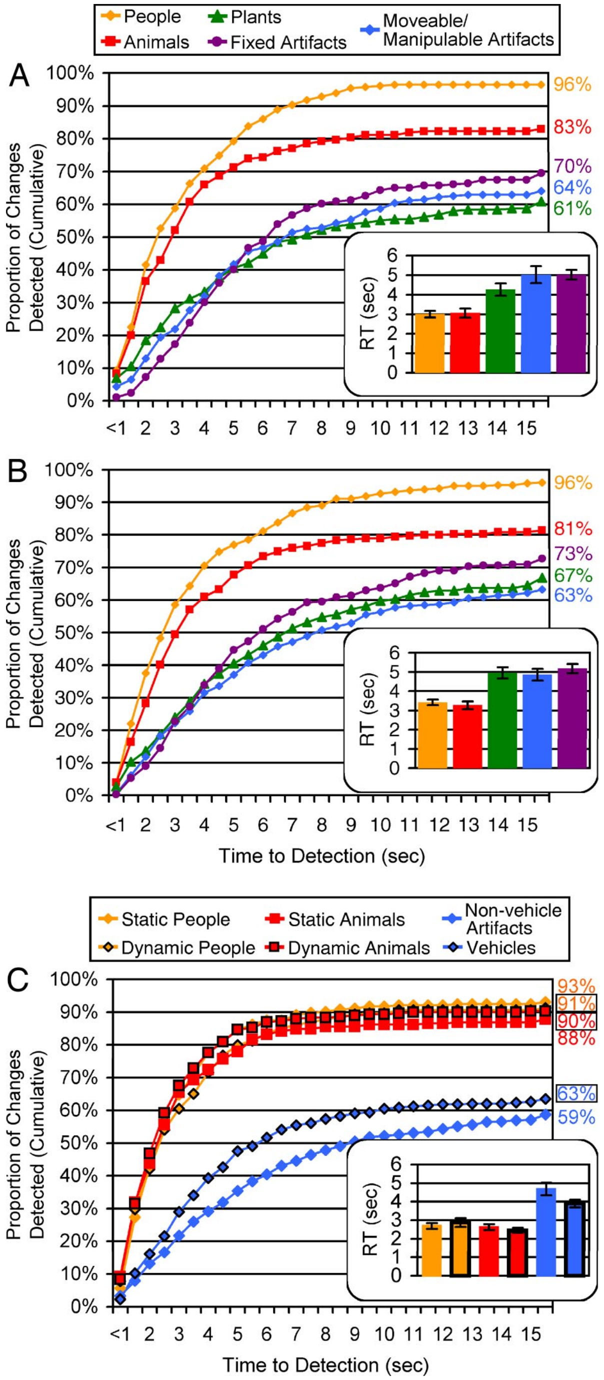


Fig. 3. Changes to animals and people are detected faster and more accurately than changes to plants and artifacts. Graphs show proportion of changes detected as a function of time and semantic category. (*Inset*) Mean RT for each category (people, animals, plants, moveable/manipulable artifacts, and fixed artifacts). (*A*) Results for Exp 1. Animate targets: RT $M = 3{,}034$ msec (SD, 882), hit rate $M = 89.8\%$ (SD, 7.4). Inanimate targets: RT $M = 4{,}772$ msec (SD, 1,404), hit rate $M = 64.9\%$ (SD, 15.7). (*B*) Results for Exp 2. Animate targets: RT $M = 3{,}346$ (SD, 893), hit rate $M = 88.7\%$ (SD, 8.0). Inanimate RT $M = 4{,}996$ (SD, 1,284), hit rate $M = 67.5\%$ (SD, 16.5). (*C*) Results for Exp 5. RT: animate $M = 2{,}661$ msec (SD, 770). Hit rate, animate vs. vehicle: 90.6% (SD, 7.8) vs. 63.5% (SD, 18.8), $P = 10^{-15}$.

group of subjects rated how interesting each target was (*SI Appendix 1*), interest ratings correlated with animacy ($r = 0.60$, $P = 10^{-7}$). But does this explain the animate attention effect?

No. Although they were correlated with animacy, interest ratings do not predict RTs once one statistically controls for whether the target is animate (partial $r = -0.16$, $P = 0.20$). In contrast, animate targets elicit faster RTs, even after controlling for interest ratings (partial $r = -0.41$, $P = 0.001$; see *SI Appendix 1.2*). The same result holds for hit rates (animacy, partial $r = 0.37$, $P = 0.002$; interest, partial $r = 0.064$, $P = 0.60$). Thus, the animacy bias was not a side effect of animals and people being more "interesting" targets, as judged by deliberative processes. Fast, accurate change detection of animates results from a category-driven process: animacy, not interest, predicts change detection efficiency.

Is Preferential Attention to Animates a Side Effect of Differences in Lower-Level Visual Characteristics? Does the animate attention effect found in Exp 1 and 2 reflect nothing more than a confound in the stimuli? Given that lower-level features (e.g., color, luminance) can affect attention in simple visual arrays (19) and more natural and complex scenes (20), it is important to eliminate the hypothesis that, in the particular stimuli we used, these features differed for animate vs. inanimate targets.

Target luminance, size (pixels), and eccentricity were entered into a multiple regression model; none predicted RT or accuracy, either across or within domains (P values range from 0.2 to 0.9). To eliminate counterhypotheses invoking any potential lower-level feature(s), Exp 3 and 4 were conducted.

Inverting photos perfectly preserves their lower-level stimulus properties but makes identifying the semantic category to which a target belongs more difficult (8, 18, 21). Further, scene inversion sizably reduces the detection of changes to high- relative to low-interest items (18) (but see ref. 22). If lower-level properties are causing the animate attention advantage, then it should appear even when photos are inverted. In contrast, if the attentional bias is category-driven, then manipulations that interfere with categorization but preserve lower-level perceptual features should eliminate the animate change-detection advantage.

Exp 3 was identical to Exp 1 and 2, except the photos were inverted. The procedure and stimuli for Exp 4 were also the same, except target category identification was disrupted not by inverting but by blurring each photo with a Gaussian blurring function in Adobe Photoshop (see *SI Appendix 2*). This preserves many lower-level characteristics (although not as perfectly as inversion) but disrupts object recognition more than inversion does. Both manipulations were used, because each has advantages that the other lacks. Each method succeeded in disrupting recognition; compared with Exp 1 and 2, RTs were slower in Exp 3 and 4, and accuracy was worse overall in Exp 4 (*SI Appendix 1.3*). If lower-level characteristics were causing the animate attention effect, then it should still appear in Exp 3 and 4. It did not (see Fig. 4).

Specifically, inverting scenes eliminated the animate advantage in detection speed ($P = 0.25$). Changes to inverted people, animals, fixed artifacts, and plants elicited comparable mean detection times (Fig. 4*A*; *SI Appendix 1.4*). When inverted, accuracy was comparable for fixed artifacts, plants, people, and animals. (Compared with other inanimate targets, accuracy for inverted moveable artifacts was disproportionately low, a pattern not seen in the upright conditions; *SI Appendix 1.5*). This is in contrast to the pattern for upright scenes, where animate beings showed a consistent speed and accuracy advantage compared with all inanimate categories.

Blurring upright scenes also eliminated the animate advantage in detection speed ($P = 0.17$). There was no animate advantage in accuracy either. In fact, the reverse was true: in the blur condition, accuracy was greater for inanimate objects (Fig. 4*B*; *SI Appendix 1.6*).

Inversion and blurring disrupt recognition, which is necessary for targets to be identified as animate vs. inanimate, while

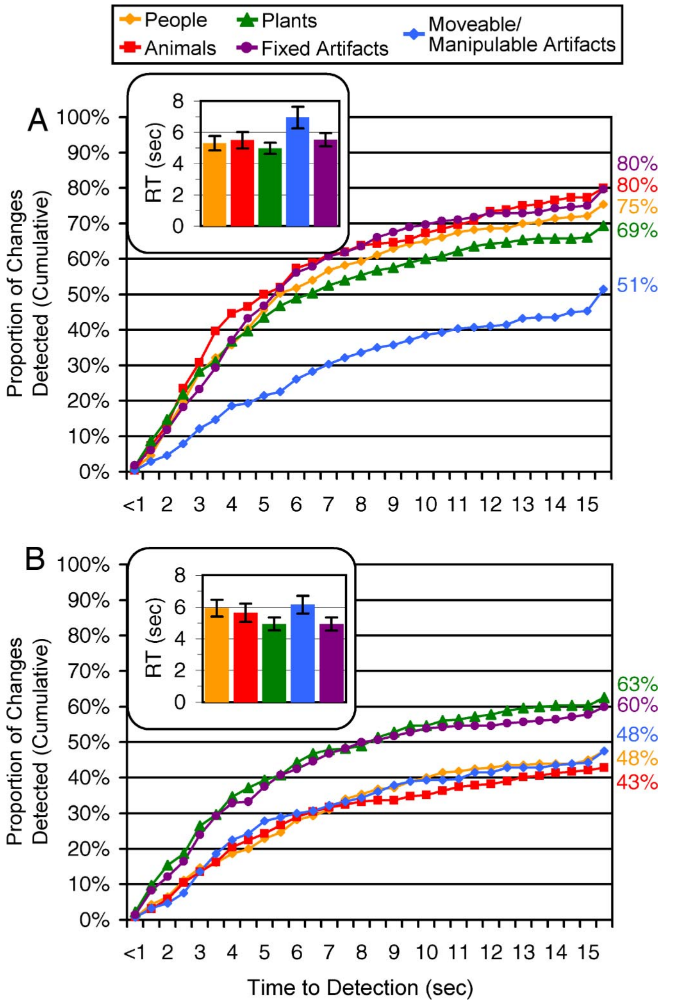


Fig. 4. Disrupting recognition eliminates the advantage of animates in change detection, showing that the animate advantage is driven by category, not by lower-level visual features. Graphs show proportion of changes detected as a function of time and category when recognition is disrupted. (*Inset*) Mean RT for each category. (*A*) Results for Exp 3 using inverted stimuli. RT, animate M = 5,399 (SD, 2,139), inanimate M = 5,813 (SD, 2,405). (See *SI Appendices 1.4* and *1.5*.) (*B*) Exp 4, blurred stimuli. RT, animate M = 5,792 (SD, 2,705), inanimate M = 5,337 (SD, 2,121). Accuracy; animate M = 45.2% (SD, 15.1), inanimate M = 56.7% (SD, 13.5), greater accuracy for inanimates; P = 0.0001, r = 0.67.

preserving all (inversion) or some (blurring) lower-level stimulus characteristics. That these two manipulations eliminated the animate detection advantage found in Exp 1 and 2 indicates the animate attentional advantage depends on recognition of the target's semantic category. It was not a side effect of incidental differences in contrast, visual complexity, or any other lower-level property of the stimuli used. (Additional controls show that the animate advantage also remains strong when controlling for potential differences in scene backgrounds; see *SI Appendix 1.7*).

Is Preferential Attention to Animates a Consequence of Experience with Motion? The animate monitoring hypothesis proposes that animates are attended by virtue of category-specific attentional mechanisms that are triggered by properties of animals and humans, not by mechanisms that attend to anything often seen in motion. Vehicles were chosen as a control category, because they move yet are not animals.

Vehicles are seen in motion every day, and the failure to monitor that motion has life-or-death consequences. Indeed, this expertise might give vehicles a detection advantage over other inanimate objects. But the prediction that animate inputs will be closely attended is not based on expertise-driven attention. It is based on the hypothesis that the visual system is designed to monitor animates because of their importance ancestrally. Consequently, animals and people should be monitored more closely than vehicles, despite our ontogenetic experience of vehicular motion and its importance to survival in the modern world. To test this, we conducted Exp 5, which specifically compared detection of animate targets to vehicles.

Of the artifact targets, 24 were vehicles (on roads, rivers, etc.), and 24 were artifacts that do not move on their own (e.g., lampposts, keys). To see whether there is any effect of implied motion on attention (23–25) due not to the target's category but rather to representations of motion or momentum (e.g., sitting vs. walking), half the people and half the animals were in motion, and half were not. Thus, there were static and dynamic animate targets and static (lampposts, keys) and dynamic (vehicles) inanimate targets. Otherwise, the procedure was the same as for Exp 1.

The results of Exp 5 are shown in Fig. 3*C*. Accuracy for vehicles and static artifacts was low (and comparable), with changes to vehicles detected faster than changes to static artifacts (P = 0.00072, r = 0.52). Nevertheless, changes to animals and people were detected >1 second faster than changes to vehicles, and the effect size was large, r = 0.82 (animate vs. vehicles, $P = 10^{-11}$). Even so, this underestimates the animate attentional advantage over vehicles, because accuracy for animate targets was 27% points higher than for vehicles, another large effect size, r = 0.87 [animate vs. vehicle, 90.6% (SD, 7.8) vs. 63.5% (SD, 18.8), $P = 10^{-12}$]. That is, subjects were change blind >36% of the time for vehicles but <10% of the time for animals and people. Detection of animate targets was better than vehicle targets at all time intervals, even <1 second.

Compared with vehicles, the speed and accuracy advantage for non-human animals was just as large as for people (animals vs. vehicles, RT, r = 0.80, $P = 10^{-10}$; hits, r = 0.84, $P = 10^{-14}$; people vs. vehicles, RT, r = 0.78, $P = 10^{-9}$; hits, r = 0.88, $P = 10^{-16}$). Moreover, the advantage for non-human animals remains just as large if the vehicle category is restricted to include only cars and trucks, the vehicles that subjects need to monitor most often (RT, r = 0.79, $P = 10^{-9}$; hits: r = 0.85 $P = 10^{-15}$).

To make sure these effects were not due to incidental differences in low-level visual characteristics, we conducted an inversion control for Exp 5 (analogous to Exp 3). Although there were some differences between categories on inversion, the animate attentional advantage in Exp 5 remains large and significant when these potential differences in low-level features are controlled for (RT, r = 0.74, $P = 10^{-7}$; hits, r = 0.88, $P = 10^{-12}$; *SI Appendix 1.8*). The same is true when one controls for potential differences in scene backgrounds (*SI Appendix 1.7*).

It is known that the human visual system has a bias to detect motion, and that momentum is represented even from still pictures (23–25). Are changes to animals and people detected faster and more accurately merely as a side effect of attention to objects in motion, animate or not?

No. For the animate monitoring effect to be a side effect of motion detection, there would have to be a CD advantage for targets in motion over stationary ones, even for the categories animal and person. Fig. 3*C* shows this was not the case; for animals and people, CD was just as fast and accurate when their pose was stationary as when it was dynamic (stationary vs. dynamic; hit RT means 2,660 msec (SD, 968) vs. 2,661 (SD, 1,142); hit rates, 91% for both). Thus implied motion does not cause a category-independent increase in attentional monitoring.

Because there were no category-independent effects of representational momentum on change detection, such effects cannot explain the CD advantage of vehicles over static artifacts.

This suggests that the vehicle vs. static advantage was caused by greater monitoring of objects identified as vehicles (whether in motion or not).

Better change detection for non-human animals than for vehicles demonstrates a category-based dissociation between recognition and monitoring. In directed categorization tasks, the visual system can rapidly detect the presence of both animals and vehicles in a natural scene (26), even in the near absence of attention (27). But the large difference in change detection demonstrated here shows that the attentional system spontaneously monitors animals more than vehicles (or other inanimates), even when there is no instruction to do so.

Is There an Effect of Ontogenetic Expertise? The CD advantage of vehicles over other inanimate objects is consistent with a modest expertise effect, although it could also be a side effect of an animate attention bias that is weakly evoked by vehicles (people make vehicles move; psychophysically, vehicular motion exhibits the contingent reactivity of animate motion) (28). But if experience were having a major effect on incidental attention, we would see a large CD advantage for vehicles over animals. Instead, the reverse is true. There would also be a large CD advantage for humans over non-human animals, a prediction that is also falsified.

In modern environments, encounters with other humans are more frequent and have greater consequences than encounters with non-human animals. So how much (or little) does ontogenetic expertise with humans promote change detection, compared with non-human animals? The curves for animals and humans are almost identical in Exp 5 (Fig. 3*C*), and they track each other closely for time intervals <3–4 seconds in Exp 1 and its replication (Exp 2). More specifically, in Exp 1, 2, and 5, there was no speed advantage for humans over animals (animals vs. humans, mean RT for hits, $P = 0.83, 0.46, 0.07$; animals were faster). Accuracy was the same in Exp 5 ($P = 0.07$) but higher for humans than for animals in Exp 1 and its replication (Exp 1, $P = 0.0003, r = 0.61$; Exp 2, $P = 10^{-7}, r = 0.76$).

Close attention to non-human animals makes sense in ancestral environments but not in the ontogenetic environment experienced by our subjects. Moreover, subjects are visually trained on the human species many orders of magnitude more than on any other species. If expertise acquisition was a function of frequency of exposure and stimulus importance, then change detection for human targets should be orders of magnitude better than for non-human animal targets. Yet there was no speed advantage for detecting changes to humans, and a lifetime of exposure to humans led only to an inconsistent advantage in accuracy: more changes were detected when the target was a person than an animal in Exp 1 and its replication but not in Exp 5. The limited differences in outcome compared with massive differences in training indicate that other causes are at play aside from, or in addition to, training. These results, like the animal–vehicle difference, call into serious question ontogenetic explanations that invoke only domain-general expertise learning.

Conclusion

Changes to animals, whether human or non-human, were detected more quickly and reliably than changes to vehicles, buildings, plants, or tools. Better change detection for non-human animals than for vehicles reveals a monitoring system better tuned to ancestral than to modern priorities. The ability to quickly detect changes in the state and location of vehicles on the highway has life-or-death consequences and is a highly trained ability; indeed, driving provides training with feedback, a situation that should promote the development of expertise-derived selection criteria. Yet subjects were better at detecting changes to non-human animals, an ability that had life-or-death consequences for our hunter–gatherer ancestors but is merely a distraction in modern cities and suburbs. This speaks to the origin of the selection criteria that created the animate monitoring bias.

The selection criteria responsible were not goal-derived: the only instructed goal was to detect changes (of any kind), and there was nothing in the structure of the task to make animals more task-relevant than inanimate objects (if anything, the reverse was true: there were more changes to inanimates than to animates). Nor were they expertise-derived: in the modern world, detecting changes in animals is an inconsequential and untrained ability compared with detecting changes in vehicles. Taken together, the results herein implicate a visual monitoring system equipped with ancestrally derived animal-specific selection criteria. This domain-specific subsystem within visual attention appears well designed for solving an ancient adaptive problem: detecting the presence of human and non-human animals and monitoring them for changes in their state and location.

Materials and Methods

Five CD experiments were conducted, each involving a different set of subjects (*SI Appendix 1*). The 70 scenes used in Exp 1–4 are shown in *SI Appendices 3–7*. The 96 scenes used in Exp 5 are shown in *SI Appendices 8–11*; there were 48 with artifact targets and 48 with animate targets (24 people and 24 animals). Of the artifact targets, 24 were vehicles, and 24 were artifacts that do not move on their own.

We thank Max Krasnow, Christina Larson, June Betancourt, Jack Loomis, and Howard Waldow and appreciate the support of the University of California, Santa Barbara (UCSB) Academic Senate (L.C.); UCSB Graduate Division (J.N.); National Institute of Mental Health National Research Service Award (F32 MH076495-02, to J.N.); and National Institutes of Health Director's Pioneer Award (to L.C.).

1. Cosmides L, Tooby J (2000) in *Metarepresentations: A Multidisciplinary Perspective*, ed Sperber D (Oxford Univ Press, New York), pp 53–115.
2. Tooby J, DeVore I (1987) in *Primate Models of Hominid Behavior*, ed Kinzey W (SUNY Press, New York), pp 183–237.
3. Shinoda H, Hayhoe M, Shrivastava A (2001) *Vision Res* 41:3535–3545.
4. Werner S, Thies B (2000) *Visual Cognit* 7:163–173.
5. Myles-Worsley M, Johnston W, Simons M (1988) *J Exp Psychol Learn Mem Cognit* 14:553–557.
6. Chun M, Jiang Y (1998) *Cognit Psychol* 36:28–71.
7. Mack A, Rock I (1998) *Inattentional Blindness* (MIT Press, Cambridge, MA).
8. Ro T, Russell C, Lavie N (2001) *Psychol Sci* 12:94–99.
9. Friesen C, Kingstone A (1998) *Psychon Bull Rev* 5:490–495.
10. Langton S, Bruce V (2000) *J Exp Hum Percept Perform* 26:747–757.
11. Downing P, Bray D, Rogers J, Childs C (2004) *Cognition* 93:B27–B38.
12. Orians G, Heerwagen J (1992) in *The Adapted Mind*, eds Barkow J, Cosmides L, Tooby J (Oxford Univ Press, New York), pp 555–579.
13. Grimes J (1996) in *Vancouver Studies in Cognitive Science*, ed Akins K (Oxford Univ Press, New York), Vol 5, pp 89–110.
14. Rensink R, O'Regan J, Clark J (1997) *Psychol Sci* 8:368–373.
15. Shapiro K (2000) *Visual Cognit* 7:83–91.
16. Caramazza A (2000) in *The New Cognitive Neurosciences*, ed Gazzaniga M (MIT Press, Cambridge, MA), pp 1199–1210.
17. Caramazza A, Shelton J (1998) *J Cognit Neurosci* 10:1–34.
18. Kelley T, Chun M, Chua K (2003) *J Vision* 2:1–5.
19. Turatto M, Galfano G (2002) *Vision Res* 40:1639–1644.
20. Parkhurst D, Law K, Niebur E (2002) *Vision Res* 42:107–123.
21. Rock I (1974) *Sci Am* 230:78–85.
22. Shore D, Klein R (2000) *J Gen Psychol* 127:27–43.
23. Freyd J (1983) *Percept Psychophys* 33:575–581.
24. Kourtzi Z, Kanwisher N (2000) *J Cognit Neurosci* 12:48–55.
25. Senior C, Barnes J, Giampietroc V, Simmons A, Bullmore E, Brammer M, David A (2000) *Curr Biol* 10:16–22.
26. Van Rullen R, Thorpe S (2001) *Perception* 30:655–668.
27. Li F, VanRullen R, Koch C, Perona P (2002) *Proc Natl Acad Sci USA* 99:9596–9601.
28. Blakemore S, Boyer P, Pachot-Clouard M, Meltzoff A, Segetbarth C, Decety J (2003) *Cereb Cortex* 13:837–844.

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.

<u>Results for Exps 1 and 2</u>. 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, *P*s = 0.12, 0.22; for interesting, *P*s = 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.

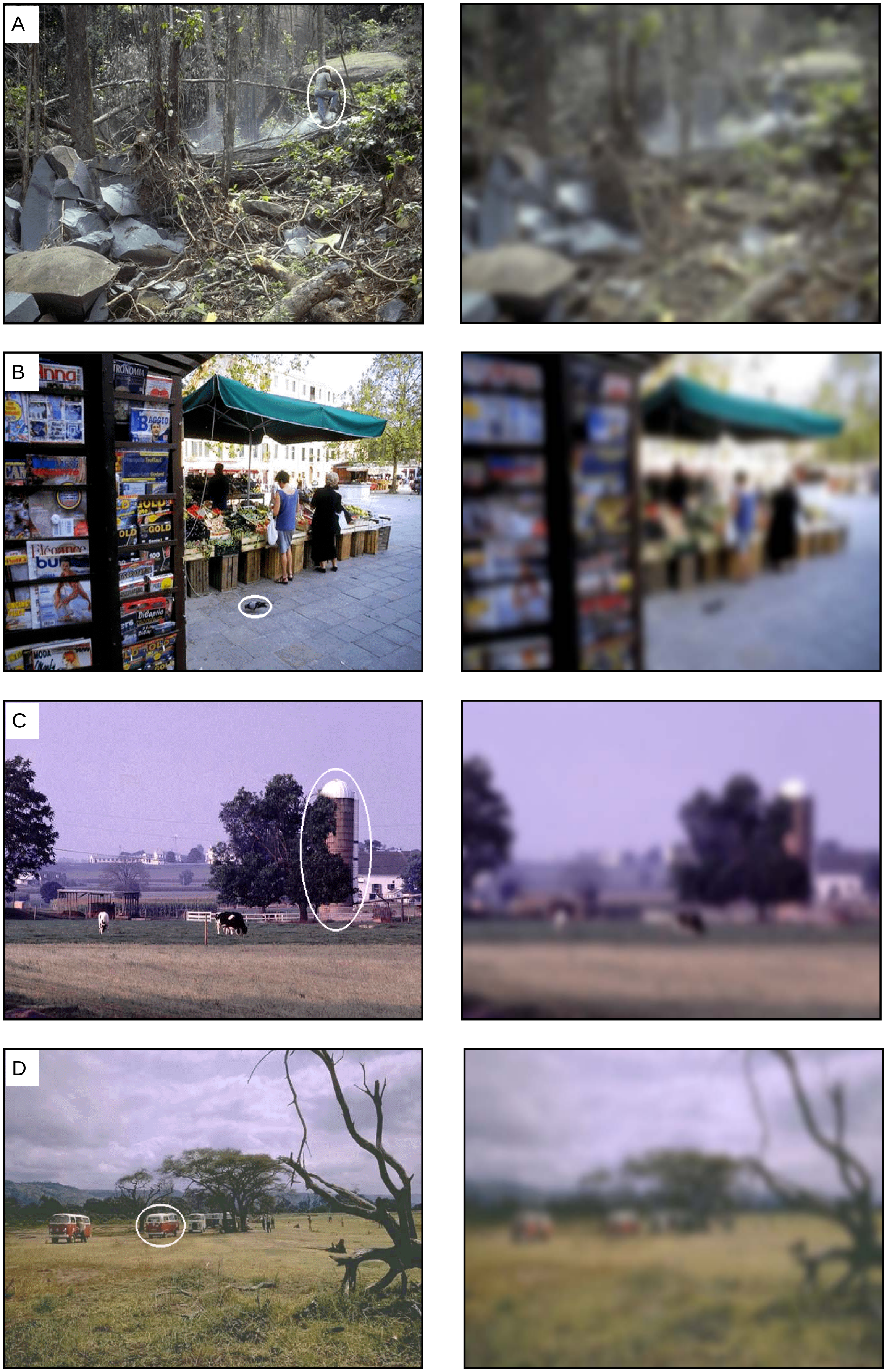
Nevertheless, the animate attentional advantage remained large and significant even after the correction for low level features was applied to the results of Exp 5. Changes to animate targets were detected more than a second faster than changes to inanimate targets (2,717 ms vs. 3,978 ms, $r = 0.74$, $P = 10^{-7}$), 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.

Sample Stimuli After Gaussian Blurring



E
F
THE WORLD