Humans are Detected More Efficiently than Machines in the Context of Natural Scenes

KATJA M. MAYER[1,2]* *Max Planck Institute for Human Cognitive and Brain Sciences*

QUOC C. VUONG *Newcastle University*

IAN M. THORNTON *University of Malta*

Abstract: In the context of natural scenes, we recently showed that detecting humans among machine distractors is more efficient than detecting machines among human distractors (Mayer, Vuong, & Thornton, 2015). We concluded that the attentional system is tuned to efficiently process human form and motion. However, our results are also consistent with the possibility that discarding machine distractors is more efficient than discarding human distractors. In the present study, we replicated our previous visual search experiment but this time embedded targets amongst the same type of distractors; namely scenes displaying natural motion (e.g., billowing clouds, trees moving in the wind). Detecting humans among natural motion was more efficient than detecting machines among the same distractors as reflected in shallower search slopes, smaller intercepts, shorter first fixation durations on targets, and higher percentages of first fixations on targets. These findings are in line with efficient detection of human targets but not with efficient discarding of machine distractors.

Key words: biological motion, eye-tracking, natural scenes.

Recently, we explored whether human bodies and actions automatically attract attention (Mayer, Vuong, & Thornton, 2015). We used a standard visual search paradigm (Eckstein, 2011; Kristjánsson, 2015; Nakayama & Martini, 2011; Wolfe, 2010, 2016) to compare the detection of video clips and images of humans to other, non-animate object categories. We were particularly interested in whether search for human targets exhibited "pop-out," a pattern in which search times are invariant to the number of distracting items in the search array, suggestive of pre-attentive, parallel processing (Treisman & Gelade, 1980; Wolfe, 2003). Several lines of previous evidence have suggested that human form and motion could attract attention in this way (New, Cosmides, & Tooby, 2007; Pratt, Radulescu, Guo, &

Abrams, 2010; Shi, Weng, He, & Jiang, 2010; Thornton & Vuong, 2004; Troje & Westhoff, 2006). However, we found no evidence for pop-out, either in search times or eye-movement parameters (Mayer et al., 2015).

Even though human form and motion did not pop-out, our data suggested that human targets were detected more efficiently than the mechanical targets that served as our main comparison category. In our experimental design, we used separate blocks in which humans were targets amongst machine distractors or machines were the targets amongst human distractors. The appearance of "search asymmetries" (Rosenholtz, 2001; Treisman & Souther, 1985; Wolfe, 2001) in both search slopes and eye movements led us to conclude that there was an attentional advantage in the efficiency of processing human form and motion.

However, as pointed out by a colleague in response to our paper (Wolfe, J. M., personal communication, 2015, October 23), there is an alternative interpretation for this search asymmetry: Our pattern of data is also consistent with the possibility that machine distractors could be recognized and discarded more quickly than human distractors. That is, when performing a serial search through an array of items (i.e., where each item in the search array is processed separately), most of the decisions will involve the distractor category, as search is terminated when the target is located. Thus, if it is easier to detect machines and move on, then this could also explain our previous results. An alternative way to think about this possibility would be if each human distractor "held" attention a little longer than each machine distractor. This could also lead to apparently less efficient search for machine targets. In line with this personal communication, a number of other sources have also suggested that the nature of distractor items can play a vital role in determining patterns of search asymmetries (Rauschenberger & Yantis, 2006; Treisman & Souther, 1985; Wolfe, 2001, 2014).

The purpose of the current brief report was to test the relative efficiency of search for humans versus search for machines in a different way by introducing a third, neutral distractor category. We used the same human and machine target videos as in our previous study and an identical experimental design, except that all distractor items were now taken from a collection of natural, outdoor scenes containing "natural motions." As in our previous study, observers searched for scenes containing a human or machine target, but this time amidst scenes containing natural motion while their eye movements were tracked. Our main question was whether human targets would still be found more efficiently than machine targets, thus supporting our previous conclusion. If visual search for human targets is more efficient than for machine targets irrespective of the distractors, we would expect to replicate our previous results. Specifically, we would expect shallower search slopes and smaller intercepts for human compared to machine targets. With respect to eye-movements, we would expect shorter fixations on human targets than on machine targets (i.e., more efficient processing of humans), and higher percentages of fixations to be initially drawn to human rather than to machine targets (i.e., humans are more likely to attract attention).

Methods

Participants. Nine participants recruited from the wider Newcastle University community completed the experiment either in return for course credit or on a voluntary basis (three females, mean age: $M = 24.7$ years, $SE = 3.3$ years). The experiment was conducted in accordance with the Declaration of Helsinki and ethical approval was provided by the ethics committee of Newcastle University. Participants gave written consent prior to the experiment. They were informed about the procedure but naive to specific hypotheses.

Stimuli and apparatus. The stimuli consisted of 1.8-s video clips (25 frames/s, 128 pixel × 96 pixel grayscale images,

5.4° × 4.1° visual angle). There were three categories of clips with eight scenes in each category: (a) human motions (kicking a football, performing a cartwheel, stretching arms, jumping jacks, doing the dishes, rolling on the floor, walking down the stairs, swinging a bat); (b) mechanical motions (pedals of a bike moving, moving carousel, cars in a street, crane transporting a bar, truck unloading stones, back-and-forth action of an industrial sawing machine, large wheel turning, machine spinning to roll up a rope); and (c) "natural motions" (clouds moving, trees blowing in the wind, a waterfall, flames of a big fire moving, strong rain fall, a river flowing, water running between rocks, tornado blowing sand). The human and the mechanical motions were identical to the video clips used in Mayer et al. (2015). There were never any objects from another category in the videos. Stimuli were taken from films and documentaries or acquired with a camcorder. All videos can be viewed at https://www.staff.ncl.ac.uk/q.c.vuong/gifs/MayerVuongThornton.html (the animated gifs are ordered according to the descriptions above). Please note that these versions of the clips are for illustration only and do not reflect the actual viewing parameters used in the experiment.

In our previous report, we simulated model observers that searched the arrays for human targets based on saliency defined by the combination of luminance, edge orientation and average amount of image motion (Mayer et al., 2015). This simulation suggested that human performance was not entirely based on these low-level features. As we changed the distractor set, here we only report the average image motion in each category; this image motion serves as a relative measure of the average speed per category. For each video, we computed the average image motion across the frames in that video. We then averaged the image motion across the eight videos in each category (humans, machines, and natural motion) and conducted t-tests to determine whether the amount of motion differed between categories. Briefly, we computed the average image motion as follows. First, for

each video we took sequential pairs of frames (1–2, 2–3, …) and computed the optic flow using the Lucas-Kanade algorithm (Lucas & Kanade, 1981) implemented in Piotr Dollar's image-processing toolbox (https://pdollar.github.io/toolbox/). The optic flow algorithm estimates the displacement of each pixel from Frame N to Frame $N + 1$, and provides a vector indicating the direction and magnitude of estimated motion displacement (with subpixel resolution). Second, we averaged the magnitudes at each pixel (i.e., the length of the vector at that pixel) and across all frame pairs to derive a single estimated image motion. This value was normalized to be within 0 and 1 (arbitrary units) so that we could average across videos. We found that the average amount of image motion of videos displaying humans and videos displaying machines did not differ (humans: $M = 0.18$ pixels/frame, $SE = 0.03$ pixels/frame; machines: $M = 0.15$ pixels/frame, $SE = 0.03$ pixels/frame; $t(14) = 0.79$, $p > .44$). The average amount of image motion of videos displaying natural motion differed from the videos displaying machines (natural motion: $M = 0.26$ pixels/frame, $SE = 0.02$ pixels/frame; $t(14) = 2.59$, $p = .021$) and marginally from the videos displaying humans ($t(14) = 2.12$, $p = .052$).

The setup was identical to that used in our previous study (Mayer et al., 2015). Participants sat in front of a Sony Trinitron CRT monitor (100 Hz refresh rate, 1,024 pixel × 768 pixel screen resolution) with their head constrained by a chin rest. The distance to the monitor was approximately 50 cm. Their right eye was tracked using a Cambridge Research System eye-tracker (50-Hz sampling rate, 0.1° spatial resolution). Stimulus display, eye-tracking, and response collection were controlled by a Windows PC, running Matlab with custom scripts written using the Psychophysics Toolbox extensions (Brainard, 1997; Kleiner et al., 2007; Pelli, 1997).

Design and procedure. All aspects of the design and procedure were identical to our previous study (Mayer et al., 2015) except for the distractor scenes. In separate blocks,

participants searched for a scene from the target category (i.e., human or machine) amidst distractor scenes (i.e., always natural motions). There were three within-subject factors: target type (human, machine), trial type (target present, target absent), and set size (two, four, six, and eight scenes in the search array). The presentation of the target category was blocked, and the order was counterbalanced across participants.

Prior to the experiment, participants were presented with all of the videos to ensure that they were familiar with the target and distractor categories. For each of the three categories of videos, they saw all eight scenes in a 2 rows × 4 columns array, and wrote a brief description of each scene on a piece of paper. The familiarization phase took approximately 7 min.

Each trial began with a white fixation cross at the center of a gray background, which remained visible throughout the trial. One second after fixation onset, the search array appeared and remained on the screen until participants responded. The videos were evenly distributed on an invisible circle with a radius of 300 pixels (12.5°) from the center of the screen with a random starting orientation on each trial (with 0° being the top of the screen). Each video began at a randomly selected frame and was repeated in a continuous loop. Participants used assigned keys on a standard USB keyboard to indicate whether a scene from the target category was present or absent. Both speed and accuracy were encouraged when responding. Errors were signaled by a 500-ms 1,500-Hz tone. Following each response there was a 500-ms blank inter-trial interval. Eye-tracking began immediately at the start of each trial.

For each target category, participants were tested with 256 trials divided into four blocks of 64 trials. Within each of these blocks, the eight scenes from the target category were shown once at each set size on present trials and distractors were randomly sampled from the eight scenes containing natural motion. Aside from these constraints, trial order was completely randomized and there were equal

numbers of present and absent trials. There was a self-timed break between blocks. We calibrated the eye-tracker before each target category block. The entire experiment took about 40 min.

Results

Accuracy was high in all experimental conditions (>95%) and will therefore not be discussed further. Median search times, search slopes, intercepts, and fixation data from correct trials are provided in Tables 1–3, and are summarized in Figure 1. Further details of the data analyses can be found in Mayer et al. (2015).

We submitted search slopes and intercepts to a 2 target type (human, machine) × 2 trial type (absent, present) repeated-measures analysis of variance (ANOVA). Search slopes were shallower for human targets compared to machine targets (humans: $M = 31$ ms/video, $SE = 11$ ms/video, 95% confidence interval (CI) [4 ms/video, 57 ms/video]; machines: $M = 49$ ms/video, $SE = 13$ ms/video, 95% CI [18 ms/video, 79 ms/video]; $F(1, 8) = 6.21$, $p = .04$, $\eta_p^2 = .44$), and shallower for present trials compared to absent trials (present: $M = 16$ ms/video, $SE = 5$ ms/video, 95% CI [4 ms/video, 28 ms/video]; absent: $M = 63$ ms/video, $SE = 19$ ms/video, 95% CI [20 ms/video, 106 ms/video]; $F(1, 8) = 11.42$, $p = .01$, $\eta_p^2 = .59$). There was no interaction between target type and trial type ($F(1, 8) = 2.36$, $p = .16$, $\eta_p^2 = .23$). Search slopes were significantly greater than zero in all conditions (1-sample t-tests, all ps < .03), indicating that there was no "pop-out" in any of the conditions.

Intercepts were smaller for human targets compared to machine targets (humans: $M = 567$ ms, $SE = 32$ ms, 95% CI [508 ms, 626 ms]; machines: $M = 625$ ms, $SE = 40$, 95% CI [508 ms, 532ms]; $F(1, 8) = 5.94$, $p = .04$, $\eta_p^2 = .43$). The main effect of trial type and the interaction between target type and trial type did not reach significance (ps > .21).

Table 1 Search times as a function of set size, target type, and trial type

Set size	Target present			Target absent		
	M [ms]	SE [ms]	95% CI [ms]	M [ms]	SE [ms]	95% CI [ms]
Human target						
2	582	27	[519, 645]	660	45	[556, 763]
4	612	37	[526, 697]	787	88	[585, 990]
6	640	41	[547, 734]	880	115	[614, 1,146]
8	654	44	[552, 755]	960	157	[597, 1,323]
Machine target						
2	657	38	[569, 746]	780	75	[608, 953]
4	681	39	[590, 772]	970	112	[712, 1,229]
6	725	56	[596, 854]	1,103	146	[765, 1,440]
8	777	67	[622, 932]	1,250	185	[823, 1,678]

Note. M = mean; SE = standard error of the mean; 95% CI = 95% confidence interval of the mean.

Table 2 Slopes and intercepts of the search functions in ms/video

	Target present			Target absent		
	M	SE	95% CI	M	SE	95% CI
Slopes						
Human target	12	4	[4, 21]	50	20	[5, 94]
Machine target	20	8	[4, 37]	77	21	[31, 124]
Intercepts						
Human target	561	26	[501, 622]	573	28	[508, 638]
Machine target	610	32	[537, 683]	640	52	[520, 760]

Note. M = mean; SE = standard error of the mean; 95% CI = 95% confidence interval of the mean.

For our eye-movement analyses, we analyzed fixation duration and the percentage of fixations for "first fixations" from trials in which participants responded correctly. For present trials, we defined "first fixation" as the first fixation to land within a 100-pixel radius (approximately 4.1°) from the center of the video containing the target (Eckstein, 2011; Mayer et al., 2015). For absent trials, we defined "first fixation" as the first fixation to land within a 100-pixel radius from the center of "any" of the videos in the search array.

A 2 (target type) × 2 (trial type) × 4 (set size) repeated-measures ANOVA revealed that first fixation durations were shorter for human compared to machine targets (humans: $M = 181$ ms, $SE = 9$ ms, 95% CI [159 ms, 202 ms]; machines: $M = 206$ ms, $SE = 9$, 95% CI [186 ms, 227 ms]; $F(1, 8) = 31.04$, $p = .001$, $\eta_p^2 = .80$), and shorter on absent compared to present trials (absent: $M = 161$ ms, $SE = 7$ ms, 95% CI [144 ms, 178 ms]; present: $M = 226$ ms, $SE = 11$ ms, 95% CI [200 ms, 252 ms]; $F(1, 8) = 105.45$, $p < .001$, $\eta_p^2 = .93$). These two factors interacted ($F(1, 8) = 19.02$, $p = .002$, $\eta_p^2 = .70$), and a post-hoc comparison indicated that there was a significant difference between human and machine targets on present trials (humans: $M = 204$ ms, $SE = 13$ ms, 95% CI [174 ms, 233 ms]; machines: $M = 248$ ms, $SE = 11$ ms, 95% CI [223 ms, 273 ms]; $t(8) = 5.38$, $p = .001$) but only marginally so on absent trials (humans: $M = 158$ ms, $SE = 7$ ms, 95% CI [142 ms, 173 ms]; machines: $M = 165$ ms, $SE = 8$ ms, 95% CI [145 ms, 184 ms]; $t(8) = 2.09$, $p = .07$). Lastly, there was a main effect of set size ($F(3, 24) = 26.83$, $p < .001$, $\eta_p^2 = .77$) and an interaction between trial type and set size

**Table 3** Fixation duration as a function of set size, target type, and trial type

	Target present			Target absent		
Set size	M [ms]	SE [ms]	95% CI [ms]	M [ms]	SE [ms]	95% CI [ms]
Human target						
2	208	14	[175, 241]	194	11	[168, 219]
4	198	16	[162, 234]	150	8	[133, 168]
6	199	13	[169, 230]	142	6	[129, 156]
8	210	12	[182, 238]	143	5	[131, 156]
Machine target						
2	263	9	[242, 285]	205	12	[177, 232]
4	243	11	[219, 268]	153	10	[131, 175]
6	236	11	[211, 260]	150	7	[133, 166]
8	249	18	[208, 291]	151	7	[136, 167]

Note. M = mean; SE = standard error of the mean; 95% CI = 95% confidence interval of the mean.

($F_{(3, 24)} = 11.96$, $p < .001$, $\eta_p^2 = .60$). No other interaction reached significance ($ps > .55$).

On present trials, we also computed the percentage of first fixations on targets as the number of first fixations that landed on a target divided by the total number of first fixations that landed on any scene for a given condition.

The percentages were submitted to a 2 (target type) × 4 (set size) repeated-measures ANOVA. Participants made significantly more first fixations on a human target compared to a machine target (humans: $M = 90\%$, $SE = 3\%$, 95% CI [83%, 98%]; machines: $M = 84\%$, $SE = 3\%$, 95% CI [77%, 91%];
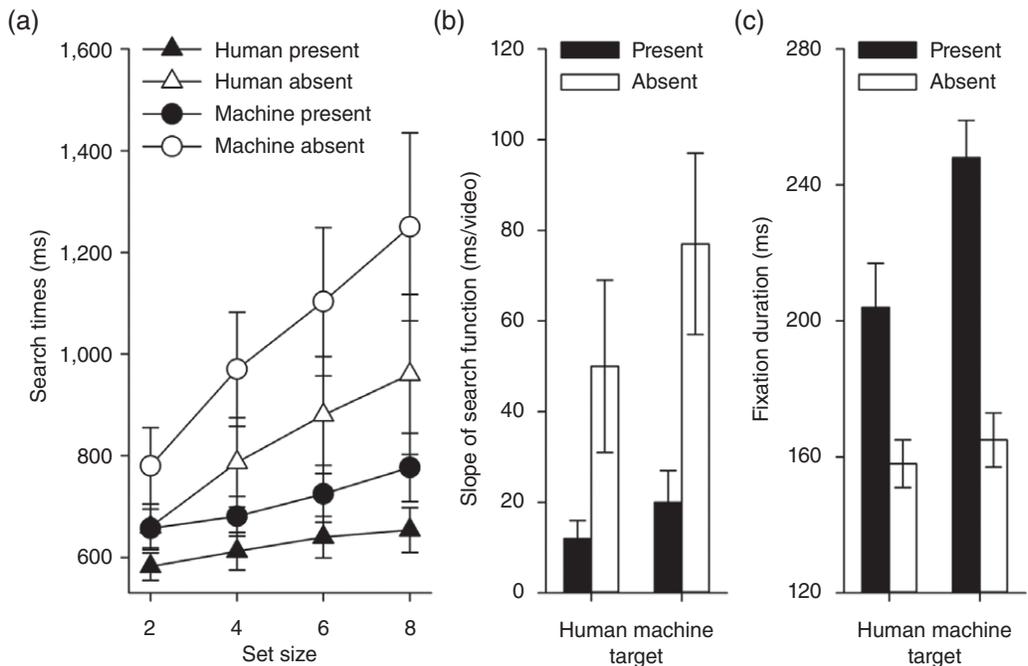
Figure 1 Results. Error bars are ±1 standard error of the means. (a) Search times as a function of target type, trial type, and set size. (b) Slopes of the linear regression lines fitted to the search times as a function of target type and trial type. (c) Fixation durations as a function of target type and trial type.

Table 4 Percentages of first fixation on target as a function of set size and target type

	Human target			Machine target		
Set size	M [%]	SE [%]	95% CI [%]	M [%]	SE [%]	95% CI [%]
2	93	3	[88, 99]	90	2	[85, 95]
4	92	5	[81, 100]	88	3	[82, 94]
6	89	4	[80, 98]	83	4	[73, 93]
8	87	5	[77, 97]	76	6	[61, 90]

Note. M = mean; *SE* = standard error of the mean; 95% CI = 95% confidence interval of the mean.

$F(1, 8) = 8.80$, $p = .02$, $\eta_p^2 = .52$; Table 4). The percentage of first fixations on a target also decreased with set size ($F(3, 24) = 4.29$; $p = .02$, $\eta_p^2 = .35$). Target type and set size did not interact ($p > .41$).

Discussion

In the present study, we investigated whether our previously reported human search advantage was due to facilitated detection of human targets or facilitated discarding of machine distractors (Mayer et al., 2015). To do this, we used a common set of "natural motion" distractor scenes with both human and machine target categories. There were three main findings. First, for both target categories, observers' search slopes were greater than zero, indicating that there was no "pop-out" for either target category. Second, observers were consistently faster at detecting human than machine targets. Lastly, observers' eye-movement data showed that first fixations on-target were shorter for human compared to machine targets and that the percentage of first fixations that landed on a target was higher for human compared to machine targets. Taken together, these findings suggest that the search advantage found for humans in our previous study is unlikely to have depended on the facilitated discarding of machine distractors.

The search pattern found for human targets in this study and in our previous study (Mayer et al., 2015) indicates a detection advantage within the human perceptual system for biological but not for mechanical objects. Elsewhere (Cavanagh, Labianca, & Thornton, 2001; Chandrasekaran, Turner, Bülthoff, & Thornton, 2010; Thornton, 2013; Thornton, Rensink, & Shiffrar, 2002) we have suggested that such an advantage could arise due to the availability of both bottom-up (Bosbach, Prinz, & Kerzel, 2004; Mather, Radford, & West, 1992; Thornton & Vuong, 2004; Troje & Westhoff, 2006) and top-down mechanisms (Bertenthal & Pinto, 1994; Bulthoff, Bulthoff, & Sinha, 1998; Thornton et al., 2002) specifically tuned for processing human form and motion. While in the current context these mechanisms are not able to automatically attract attention to human targets, they may nonetheless provide a detection advantage.

Although the current data only allow us to speculate, we would suggest that some form of top-down guidance plays a crucial role in the human advantage reported here and in our previous study (Mayer et al., 2015). As both actors and observers, we have a tremendous amount of experience with the human body and it is becoming clear that there are a variety of brain areas specifically involved with the processing of both human form (Downing, 2001; Downing & Peelen, 2011; Peelen, 2004; Schwarzlose, 2005; Vangeneugden, Peelen, Tadin, & Battelli, 2014) and human motion (Giese & Poggio, 2003; Grossman & Blake, 2001; Saygin, 2007; Thompson & Parasuraman, 2012). Cavanagh et al. (2001) suggested that this experience and specialization may have given rise to "attentional sprites," dynamic templates that can guide the processing of human targets in a top-down manner. In addition to providing an advantage in guiding attention towards a target (i.e., influencing search efficiency or slopes), such templates

could also provide a basis for more fluent perceptual processing (i.e., influencing search intercepts) by speeding decisions once a human target has been found. Of course, at least on target present trials, lower intercept times for human compared to machine targets might also reflect more basic, bottom-up perceptual differences between the categories.

Importantly, observers in our study could also deploy these putative dynamic human templates to guide visual search even when a human target is not present in the search array, particularly since they searched for human or machine targets in separate blocks. Consistent with this possibility, the search patterns on absent trials in the current study were different between the two target categories even though the search arrays were identical across both categories. Specifically, absent search slopes were 27 ms/video shallower and search intercepts were 67 ms lower in the human blocks than in the machine blocks (note that there was a main effect of target type but no significant interaction between trial type and target type for both of these measures). Again, we acknowledge that the overall speed advantage (i.e., intercept differences) in human blocks could also arise from more basic, post-search decisions processes.

We should finally note that one aspect of our target categories may have favored the deployment of top-down strategies during human search. That is, our machine targets included a range of different types of machines, with different underlying forms and, consequently, motions. In contrast, our videos displaying biological motion were all from the same basic category (i.e., human bodies) with a more constrained set of possible motions. Thus, human targets may have lower variability in their form and motion between videos. Although we did attempt to familiarize participants with videos from all categories prior to the search task, having a single exemplar type for the human category may still have afforded a top-down advantage unrelated to the fact that they were human bodies *per se*.

Related to the issue of variability across videos, the similarity between humans and

distractors and between machines and distractors may differ as a result. Duncan and Humphreys (1989) showed that both target-target and target-distractor similarity could account for a range of search efficiencies when observers searched for letters. Potential differences in these similarity relationships could provide another alternative to top-down guidance. For example, the perceived similarity between machine targets and natural motion in our experiment may be higher than between human targets and natural motion even though image motion significantly differed between videos displaying machines and videos displaying natural motion whereas an only marginally significant difference was found between image motion of videos displaying human targets and videos displaying natural motion. Therefore, participants may have taken more time to confirm the absence of a machine target because they may have been more conservative in their responses on absent trials.

In future studies, it may therefore be informative to vary the range of exemplars taken from specific mechanical and biological categories. This selection would control for the variability of form and motion between the videos in each target category. For example, we could present mechanical targets in which there are smaller differences between the exemplars (e.g., different industrial robots). Conversely, we could extend the range of the biological category to include species other than humans (e.g., different species of dogs). We can also manipulate the similarity between the non-human target videos (e.g., using exemplars from the same or different non-human category) to determine whether search efficiency for humans is due to top-down guidance *per se* or whether it may also be driven by stimulus similarity (Duncan & Humphreys, 1989).

Conclusion

The present results support the conclusion of our previous study (Mayer et al., 2015) in that

observers searched for human targets more quickly and more efficiently than they did for machine targets. As we used the same distractors for both target categories in the current work, these results help to rule out the possibility that machine distractors could be recognized and discarded more quickly than human distractors during visual search. That is, faster and more efficient search for humans appears to be due to an attentional advantage for detecting human targets rather than an advantage for discarding machine distractors.

References

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, *5*, 221–224. doi: 10.1111/j.1467-9280.1994.tb00504.x

Bosbach, S., Prinz, W., & Kerzel, D. (2004). A Simon effect with stationary moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 39–55. doi: 10.1037/0096-1523.30.1.39

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436.

Bulthoff, I., Bulthoff, H. H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, *1*, 254–257. doi: 10.1038/699

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition*, *80*, 47–60.

Chandrasekaran, C., Turner, L., Bülthoff, H. H., & Thornton, I. M. (2010). Attentional networks and biological motion. *Psihologija*, *43*, 5–20. doi: 10.2298/PSI1001005C

Downing, P. E. (2001). A cortical area selective for visual processing of the human body. *Science*, *293*, 2470–2473. doi: 10.1126/science.1063414

Downing, P. E., & Peelen, M. V. (2011). The role of occipitotemporal body-selective regions in person perception. *Cognitive Neuroscience*, *2*, 186–203. doi: 10.1080/17588928.2011.582945

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, *96*, 433–458. doi: 10.1037/0033-295X.96.3.433

Eckstein, M. P. (2011). Visual search: A retrospective. *Journal of Vision*, *11*, 14. doi: 10.1167/11.5.14

Giese, M. A., & Poggio, T. (2003). Cognitive neuroscience: Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*, 179–192. doi: 10.1038/nrn1057

Grossman, E. D., & Blake, R. (2001). Brain activity evoked by inverted and imagined biological motion. *Vision Research*, *41*, 1475–1482. doi: 10.1016/S0042-6989(00)00317-5

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3? *Perception*, *36*, 1–16.

Kristjánsson, Á. (2015). Reconsidering visual search. *i-Perception*, *6*, 2041669515614670. doi: 10.1177/2041669515614670

Lucas, B. D., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. *Proceedings of the 7th International Joint Conferences on Artificial Intelligence* (Vancouver, Canada), 674–679.

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London Series B: Biological Sciences*, *249*, 149–155.

Mayer, K. M., Vuong, Q. C., & Thornton, I. M. (2015). Do people "Pop Out"? *PLoS One*, *10*(10), e0139618. doi: 10.1371/journal.pone.0139618

Nakayama, K., & Martini, P. (2011). Situating visual search. *Vision Research*, *51*, 1526–1537. doi: 10.1016/j.visres.2010.09.003

New, J., Cosmides, L., & Tooby, J. (2007). Category-specific attention for animals reflects ancestral priorities, not expertise. *Proceedings of the National Academy of Sciences*, *104*, 16598–16603. doi: 10.1073/pnas.0703913104

Peelen, M. V. (2004). Selectivity for the human body in the fusiform gyrus. *Journal of Neurophysiology*, *93*, 603–608. doi: 10.1152/jn.00513.2004

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.

Pratt, J., Radulescu, P. V., Guo, R. M., & Abrams, R. A. (2010). It's alive! Animate motion captures visual attention. *Psychological Science*, *21*, 1724–1730. doi: 10.1177/0956797610387440

Rauschenberger, R., & Yantis, S. (2006). Perceptual encoding efficiency in visual search. *Journal of Experimental Psychology: General*, *135*, 116–131. doi: 10.1037/0096-3445.135.1.116

Rosenholtz, R. (2001). Search asymmetries? What search asymmetries? *Perception & Psychophysics*, *63*, 476–489.

Saygin, A. P. (2007). Superior temporal and premotor brain areas necessary for biological motion perception. *Brain*, *130*, 2452–2461. doi: 10.1093/brain/awm162

Schwarzlose, R. F. (2005). Separate face and body selectivity on the fusiform gyrus. *Journal of Neuroscience*, *25*, 11055–11059. doi: 10.1523/JNEUROSCI.2621-05.2005

Shi, J., Weng, X., He, S., & Jiang, Y. (2010). Biological motion cues trigger reflexive attentional orienting. *Cognition*, *117*, 348–354. doi: 10.1016/j.cognition.2010.09.001

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, *59*, 4–13. doi: 10.1016/j.neuroimage.2011.05.044

Thornton, I. M. (2013). Top-down versus bottom-up processing of biological motion. In K. L. Johnson & M. Shiffrar (Eds.), *People watching: Social perceptual, and neurophysiological studies of body perception* (pp. 25–43). New York: Oxford University Press.

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, *31*, 837–853. doi: 10.1068/p3072

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology*, *14*, 1084–1089. doi: 10.1016/j.cub.2004.06.025

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Treisman, A. M., & Souther, J. (1985). Search asymmetry: A diagnostic for preattentive processing of separable features. *Journal of Experimental Psychology: General*, *114*, 285–310.

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a "Life Detector"? *Current Biology*, *16*, 821–824. doi: 10.1016/j.cub.2006.03.022

Vangeneugden, J., Peelen, M. V., Tadin, D., & Battelli, L. (2014). Distinct neural mechanisms for body form and body motion discriminations. *Journal of Neuroscience*, *34*, 574–585. doi: 10.1523/JNEUROSCI.4032-13.2014

Wolfe, J. M. (2001). Asymmetries in visual search: An introduction. *Perception & Psychophysics*, *63*, 381–389.

Wolfe, J. M. (2003). Moving towards solutions to some enduring controversies in visual search. *Trends in Cognitive Sciences*, *7*, 70–76.

Wolfe, J. M. (2010). Visual search. *Current Biology*, *20*, R346–R349. doi: 10.1016/j.cub.2010.02.016

Wolfe, J. M. (2014). Approaches to visual search: Feature integration theory and guided search. In A. C. Nobre & S. Kastner (Eds.), *Oxford handbook of attention* (pp. 11–50). New York: Oxford University Press.

Wolfe, J. M. (2016). Visual search revived: The slopes are not that slippery: A reply to Kristjansson (2015). *i-Perception*, *7*. doi: 10.1177/2041669516643244