Congruency Effects with Animal and Human Target Objects

Qunjie Huang[1,a], Jan Drewes[2,b], Weina Zhu[1,c*]

[1]School of Information Science, Yunnan University, Kunming, China
[2]Institute of Brain and Psychological Sciences, Sichuan Normal University, Chengdu, China

[a]huang-qunjie@mail.ynu.edu.cn;[b]mail@jandrewes.de
[c*]Corresponding author: zhuweina.cn@gmail.com

Abstract. It is well-established that the visual system rapidly recognizes animals and animate objects. Also, objects semantically congruent with a scene are identified more swiftly than incongruent ones. However, it remains unclear which specific types of animate objects can be recognized most quickly. The aim of this study is to investigate whether humans are recognized faster than other animals. In our experiment, participants swiftly determined whether targets within congruent and incongruent scenes belonged to humans or animals using a two-alternative forced choice (2AFC) task. The stimulus set included an equal number of congruent and incongruent synthesized images. Our study produced two key findings: Firstly, a congruency effect was observed concerning the recognition of humans. Secondly, the visual system recognized complete human bodies more rapidly than animals. These findings extend our understanding of the congruency effect in animal perception and imply the potential existence of particular mechanisms that may facilitate rapid visual recognition of humans.

Keywords: visual perception, object recognition, body perception, rapid animal detection, congruency effect

1 Introduction

Uncovering the mechanisms of object recognition in the human visual system is an essential issue, and it is a well-established fact that humans can rapidly identify various objects, including animals[1,2], various types of objects[3], and faces[4], often rapidly, within a few hundred milliseconds[5], whether they are presented in isolation or within complex backgrounds. Prior research has demonstrated that animal targets elicit shorter reaction times and higher accuracy in identification compared to inanimate targets, even when placed in semantically incongruent contexts[6]. In the case of human objects, we can identify their body outlines within approximately 600 milliseconds[7]. The visual system can recognize common upright body postures in about 700 milliseconds[8]. humans in typical postures exhibit shorter reaction times and higher accuracy than atypical postures[9]. While studies have revealed an inversion effect for the human body, the visual system can still recognize inverted human bodies in approximately 918 milliseconds[10],

K. M. Salleh et al. (eds.), *Proceedings of the 2023 International Conference on Applied Psychology and Modern Education (ICAPME 2023)*, Advances in Social Science, Education and Humanities Research 803,
https://doi.org/10.2991/978-2-38476-158-6_6

suggesting that the brain can also process abnormal, inverted human information. Furthermore, human faces are rapidly detected and recognized by the visual system[11,12]. In general, humans are quickly recognized by the visual system thanks to features such as body shape, typical posture, and facial features.

The visual system automatically processes the global-level information of a complex scene before focusing on the local-level details[13]. However, there is few studies about how semantic contexts influence the perception of humans. In other words, does the congruency effect exist in the perception of humans? And whether such human figures are recognized from non-human animals.

2 Method

2.1 Participants

24 participants (15 males, 9 females) aged 18 to 27 (mean = 22.83, SD = 2.41) took part in the experiment. Two participants were excluded due to non-compliance with task instructions. All participants reported no history of mental or neurological disorders and had normal or corrected-to-normal vision.

2.2 Apparatus

The visual stimuli were presented on a 19-inch Sun monitor with a refresh rate of 120 Hz and resolution 1024×768. The procedure was programmed using MATLAB (The MathWorks, Inc., 2012).

2.3 Stimuli

We prepared 400 scene images for the experiment, each featuring a foreground object placed within a background context. We created a new dataset consisting of 200 images in the animal group (male or female lions) and 200 images in the human group (men and women). We selected upright human figures without additional actions as objects to ensure that the character's actions did not influence the experimental results.

All background images were resized to a resolution of 320 by 320 pixels. Subsequently, the objects were extracted and seamlessly integrated into their corresponding background images using Photoshop. For image composition, background images of the grassland and desert types were combined with lion images to create scenes with both semantic consistency and inconsistency (e.g., lion-grassland, lion-desert). The same rules were applied to beach and snowfield scenes, which featured human images.

Additionally, we performed object occupation balancing and image histogram equalization on the image set. We recorded the number of pixels occupied by each object in the human and animal groups and calculated their occupancy proportions. After manual adjustments, the difference in average object proportions between the animal and human groups was only 1.33%. To ensure uniform luminance and contrast across the images, we employed the SHINE toolbox[14]. see Fig.1.

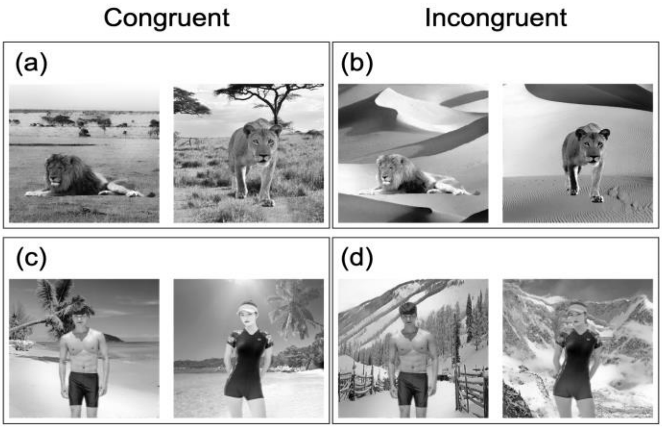

Fig. 1. Stimuli samples, animal/ human in congruent (a)(c) and incongruent (b)(d) background

2.4 Procedure

Experiment included 4 four blocks with 100 stimuli. Each trial followed a 2 (human/animal) × 2 (male/female) × 2 (congruent/incongruent) design, resulting in 50 congruent and 50 incongruent images per block. (See Fig.2 for a visual representation).

We employed the 2AFC (Two-Alternative Forced Choice) paradigm. A stimulus image was presented for 16.67 ms. Participants were instructed to classify the target (human/animal) by pressing either the left or right key on the keyboard as quickly as possible. If a response exceeded 5 seconds, the trial was automatically skipped, and no data was recorded for that trial. Following the participant's selection, a backward mask was presented for 500 ms to reduce a possible afterimage effect. If participants' accuracy fell below 90%, their data were excluded from further analysis.

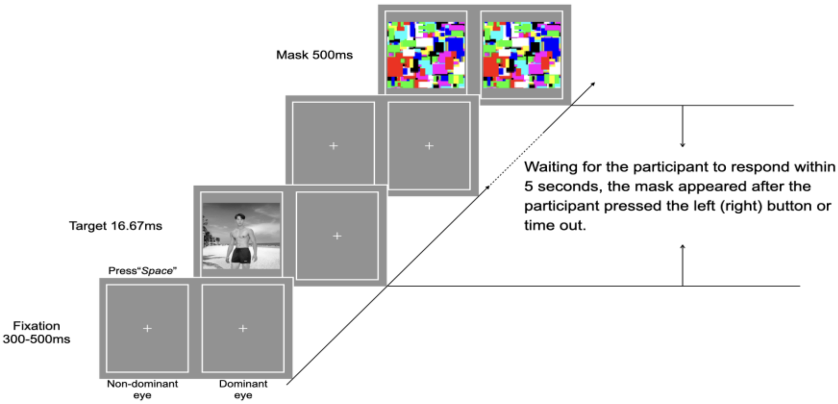

Fig. 2. Paradigm of the experiment

3 Results

In this study, we conducted a two-way ANOVA with object types (animal and human) and congruency types (congruent and incongruent) as the two independent variables to assess potential differences between these factors. We observed a significant main effect of object type ($F(1,21) = 8.322$, $p=0.009 < 0.05$), indicating that reaction times varied with changes in object type. Specifically, participants recognized human bodies faster than animal objects. Additionally, we found a significant effect of congruency ($F(1,21) = 26.482$, $p=0.000 < 0.05$), suggesting that participants performed better when the object was semantically congruent with the background. (See Fig.3).

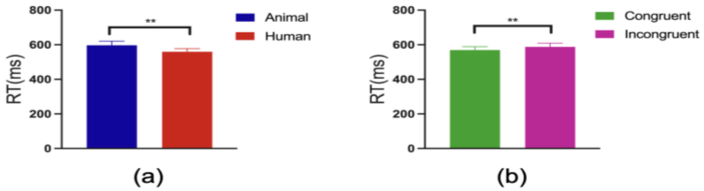

Fig. 3. (a) represents the mean reaction times for the animal and human groups, while (b) displays the mean reaction times for the congruent and incongruent groups in the experiment

Furthermore, we used an ANOVA to investigate how cognitive ability is reflected in the differences in recognition accuracy and reaction time. In the experiment, the human group had higher accuracy than the animal group (ranging from 94.9% ± 3.3 to 96.5% ± 2.5), with the highest mean accuracy in the group where the human was congruent with the background information (96.5% ± 2.5) and the lowest mean accuracy in the group where the animal was incongruent with the background (94.9% ± 3.3). In addition, we evaluated the reaction time of participants for recognition of the animal and human groups and found that the animal congruent group showed a shorter reaction time ($F(1,21)=4.714$, $P=0.042<0.05$) than the animal incongruent group, and similarly, the human congruent group had a faster reaction time ($F(1,21)=9.112$, $P=0.007<0.05$) than the human incongruent group. (Fig.4).

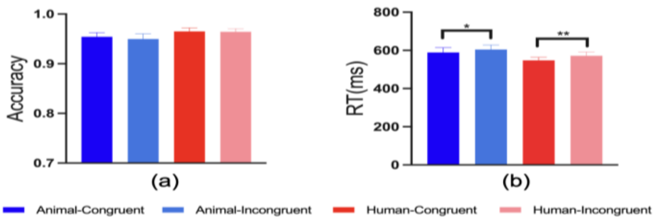

Fig. 4. (a) and (b) depict the bar plots illustrating the accuracy and reaction time for the experiment. Each bar represents one of the four groups: animal congruent, animal incongruent, human congruent, and human incongruent. The accuracy values are percentages, with the mean and standard deviation indicated

4 Conclusions

In this study, we arrived at several conclusions. Firstly, there was no difference in accuracy between the congruent and incongruent groups for both humans and animals in the experiment. However, the differences in reaction time were significant. This suggests that a congruency effect exists for human objects, wherein a semantically incongruent background impairs the recognition speed of humans. Secondly, participants exhibited quicker reaction times when detecting humans than with animal objects, indicating that humans, characterized by a human face and typical posture, are recognized more rapidly than animals.

Future research should focus on two main areas. Firstly, to address limitations in our study, subsequent experiments could include various animal objects as control groups. Furthermore, investigating cognitive differences through unconscious experiments might be beneficial. Secondly, recognition performance may vary depending on the participant's familiarity with the animal object. Shape is crucial for determining animacy in human judgment[15]. The visual system might swiftly process familiar animal shapes. Thus, future research could employ familiar animal silhouettes as stimuli for comparative studies.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (61563056, 62166049).

References

1. Crouzet, S. M., Joubert, O. R., Thorpe, S. J., & Fabre-Thorpe, M. (2012). Animal detection precedes access to scene category. PLoS One, 7(12), e51471. doi:10.1371/journal. Pone. 0051471.
2. VanRullen, R., & Thorpe, S. J. (2001). Is it a bird? Is it a plane? Ultra-rapid visual categorisation of natural and artifactual objects. Perception, 30(6), 655-668. doi:10.1068/p3029.
3. Praß, M., Grimsen, C., König, M., & Fahle, M. (2013). Ultra rapid object categorization: effects of level, animacy and context. PLoS One, 8(6), e68051. doi: 10.1371/journal. pone. 0068051.
4. Hershler, O., Golan, T., Bentin, S., & Hochstein, S. (2010). The wide window of face detection. J Vis, 10(10), 21. doi:10.1167/10.10.21.
5. Potter, M. C. (2012). Recognition and memory for briefly presented scenes. Front Psychol, 3, 32. doi:10.3389/fpsyg.2012.00032.
6. Joubert, O. R., Fize, D., Rousselet, G. A., & Fabre-Thorpe, M. (2008). Early interference of context congruence on object processing in rapid visual categorization of natural scenes. J Vis, 8(13), 11 11-18. doi:10.1167/8.13.11.
7. Wittinghofer, K., de Lussanet, M. H., & Lappe, M. (2010). Category-specific interference of object recognition with biological motion perception. J Vis, 10(13), 16. doi: 10. 1167/ 10. 13.16.

8. Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. Psychological science, 14(4), 302-308. doi:10.1111/1467-9280.14431.
9. Reed, C. L., Nyberg, A. A., & Grubb, J. D. (2012). Contributions of visual and embodied expertise to body perception. Perception, 41(4), 436-446. doi:10.1068/p7029.
10. Yovel, G., Pelc, T., & Lubetzky, I. (2010). It's all in your head: Why is the body inversion effect abolished for headless bodies? Journal of Experimental Psychology: Human Perception and Performance, 36(3), 759-767. doi:10.1037/a0017451.
11. Crouzet, S. M., & Thorpe, S. J. (2011). Low-level cues and ultra-fast face detection. Front Psychol, 2, 342. doi:10.3389/fpsyg.2011.00342.
12. Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: face detection in just 100 ms. J Vis, 10(4), 16 11-17. doi:10.1167/10.4.16.
13. Furtak, M., Mudrik, L., & Bola, M. (2022). The forest, the trees, or both? Hierarchy and interactions between gist and object processing during perception of real-world scenes. Cognition, 221, 104983. doi: 10.1016/j.cognition.2021.104983.
14. Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. Behavior research methods, 42, 671-684. doi:10.3758/BRM.42.3.671
15. Lloyd-Jones, T. J., & Luckhurst, L. (2002). Outline shape is a mediator of object recognition that is particularly important for living things. Memory & Cognition, 30(4), 489-498. 10.3758/BF03194950