Check for updates

Artificial Retina-Based Metaverse with Bionic Vision Processing

Haiyang Hu, Chao Ping Chen*, Gang Li**, Seak Pang Zou*, Qiang Chu*, Baoen Han**
*Smart Display Lab, Shanghai Jiao Tong University, Shanghai, China
**School of Psychology and Neuroscience, University of Glasgow, Glasgow, United Kingdom
Email: ccp@sjtu.edu.cn

Abstract
An artificial retina featuring bionic vision processing is presented as a paradigm shift of metaverse. The main contribution is to provide a methodology to better understand the human vision and to reproduce images along the visual pathway. Simulation results include retinal images, bilateral neural images, depth maps, etc.

Author Keywords
Artificial retina; ganglion; bionic vision; implantable device; virtual reality; metaverse.

1. Introduction
Throughout the history, the great philosophers have always been pondering over three fundamental questions. What are we? Where were we from? Where will we be? In the current mainstream view, we are human, or rather, the homo sapiens. We evolved from apes, who once lived in Africa. As to the third question, there will be a lot of possibilities. In one scenario pictured by Prof. Yuval Noah Harari in his book *Homo Deus*, homo sapiens will be upgraded into cyborgs—a being with both organic and biomechatronic or bionic parts. This may sound like science fiction, but it is becoming a reality. As early as 2011, a retinal prosthesis called Argus II (Second Sight) has begun its clinical trials in Europe. On May 25, 2023, Neuralink announced that it received approval from Food and Drug Administration to put chips into the human brain. As the life expectancy continues to increase, it is inevitable that some parts of our body will fail and need replacements by the artificial organs or implants. This would disrupt many things we know today. Metaverse, for example, can be accessed from implantable devices rather than wearable devices [1–6].

Inspired by the above prophecy, we hereby envision an implantable metaverse solution, which consists of two key components, *i.e.*, epiretinal implant and bionic vision processing. In the subsequent sections, we will discuss the implant design, physiology of vision, principles of bionic vision, and substantiate our concept by reproducing the step-wise images along the visual pathway.

2. Artificial Retina
Artificial retina or retinal implant is a type of electronic implantable device for the vision restoration [7]. Depending on where the implant is located, there are three major architectures: epiretinal, subretinal, and suprachoroidal [8]. The epiretinal implants are so far the mostly practiced, for they are easier and less risky to perform the surgery. Fig. 1 is a schematic of the proposed artificial retina, where merely the ganglions, bipolar cells, photoreceptors, and pigmented epithelium of retina are shown for the sake of simplicity. This implant is mounted on top of the layer of ganglions and has an array of electrodes that can stimulate the axons of ganglions—the bundle of which forms the optic nerve—to transmit the visual information to the brain. Surprisingly, the ganglions can be divided into various subtypes of cells. To date, only three subtypes are clearly identified. Subtype 1: midget cells or P-type ganglions. They account for about 80% of ganglions and play many important roles, including the color, spatial resolution, binocular parallax-induced depth perception, *etc.* Subtype 2: parasol cells or M-type ganglions. In contrast to midget cells, parasol cells are larger in size and fewer in number (10% of ganglions). They are highly sensitive to luminance and hence associated with the flicker perception, motion perception, motion parallax-induced depth perception, *etc*. Subtype 3: photosensitive ganglion cells. They constitute 1% of ganglions and are intrinsically photosensitive due to the melanopsin inside. They contribute to the circadian rhythm and pupillary light reflex. We shall ignore the non-image-forming subtypes (*e.g.*, parasol and photosensitive ganglion cells) and limit our interest to the midget cells for the bionic vision processing.

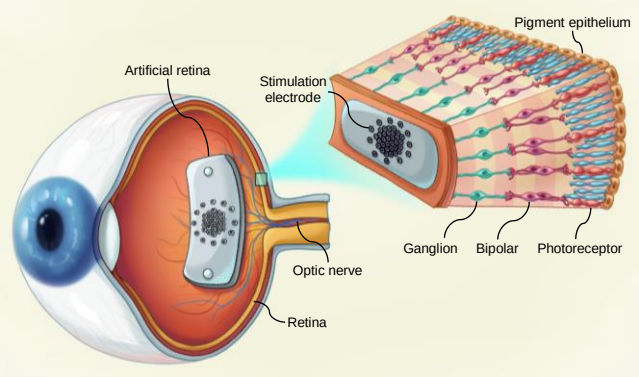

Figure 1. Schematic of the proposed artificial retina.

3. Bionic Vision Processing
Visual Pathway: To explain our bionic vision processing [9], we shall first go through the pipelines of visual pathway, which are summarized as four steps. Step 1: light travels through the cornea, anterior and posterior chambers, pupil, lens and vitreous chamber before reaching the retina. Step 2: photoreceptors of the retina convert the light into the biological signals, via the visual phototransduction. Step 3: retinal images are transmitted through the optic nerve, lateral geniculate nucleus and optic radiation before reaching the primary visual cortex, during which the visual fields will be split at the chiasm and optic radiation. Step 4a: primary visual cortex fuses two monocular images into a bilateral neural image. Step 4b: primary visual cortex detects the edge and saliency. Step 4c: primary visual cortex percepts the binocular parallax-induced depth for the overlap of binocular fields.

Foveated Blurring: Due to the uneven distribution of midget cells across the retina, the image clarity shall be proportional to the density of midget cells. We resort to an interactive software to extract the raw data of density of midget cells at different locations of retina [10]. By doing so, the contour map of midget cell density (transformed into standard visual acuity) can be depicted in Fig. 2. Blurring is implemented by applying different sizes of blurring kernels in different visual regions based on the density distribution of midget cells. The kernel sizes with respect to the offsets from the fovea are listed in Table 1. In contrast to the photoreceptor-based foveation [11], the midget-cell-based foveation is arguably more accurate at the neural level.

ISSN 0097-996X/24/5501-0545-$1.00 © 2024

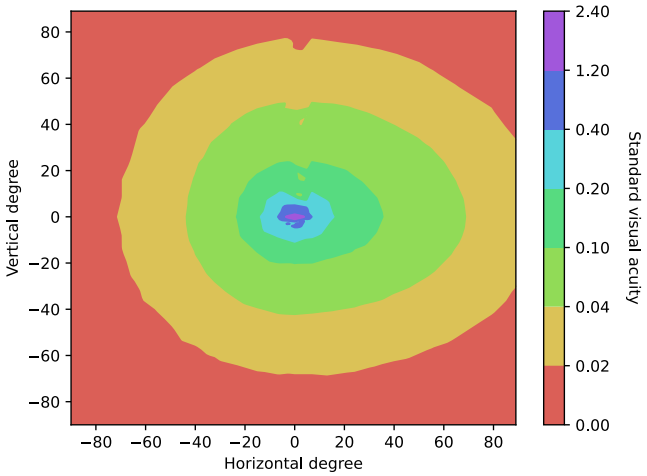

Figure 2. Contour map of the standard visual acuity.

Table 1. Kernel sizes at different locations of the visual fields.

Kernel size	Offset from fovea (°)	Offset from fovea (%)
1 × 1	25	7.45
3 × 3	28	8.40
5 × 5	31	9.51
7 × 7	34	10.62
9 × 9	37	11.89
11 × 11	41	13.63
13 × 13	43	15.37
15 × 15	48	17.59
17 × 17	52	20.13
19 × 19	56	23.45
21 × 21	61	28.53
23 × 23	65	33.91
25 × 25	70	43.42
27 × 27	76	63.55
29 × 29	81	100

Bilateral Neural Image Fusion: When the binocular visual fields are to overlap, a conflict between the left and right brains will arise. There are two ways to resolve this conflict: suppressing the image at one side of brain, or fusing left and right images. Either method is used, the phenomenon known as the eye dominance shall take effect. As shown in Fig. 3, it demonstrates the tendency of brains to prefer the visual information from the dominant eye. Statistically, the dominant eye of the majority is the right eye.

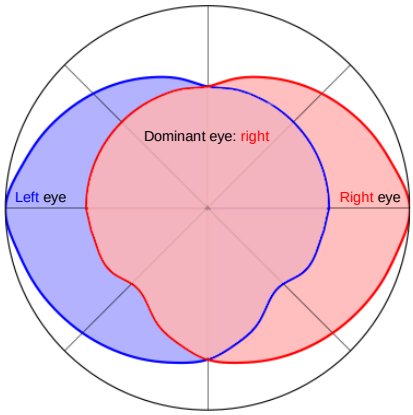

Figure 3. Right eye dominance.

Depth Perception: The cues for depth perception can be classified into two categories: binocular cues and monocular cues. The primary binocular cue is considered the binocular disparity. The depth can be inferred by

$$d = \frac{bf}{r-l} \tag{1}$$

where d is the depth of the point of interest, b the interpupillary distance, and f the focus length of the eye, l/r the bias from the center in the positive horizontal direction of the projections of the point of interest on the left/right retinal plane, as shown in Fig. 4. When the viewing or gaze directions are not parallel, it is necessary to run the epipolar correction before calculating the depth. This preprocessing aims to project the images onto a common plane, thereby simplifying the process of identifying matching points between the two input images. Another important component of the perception is the parallax estimation, which aims to find the matched pairs of points in the left and right retinal images.

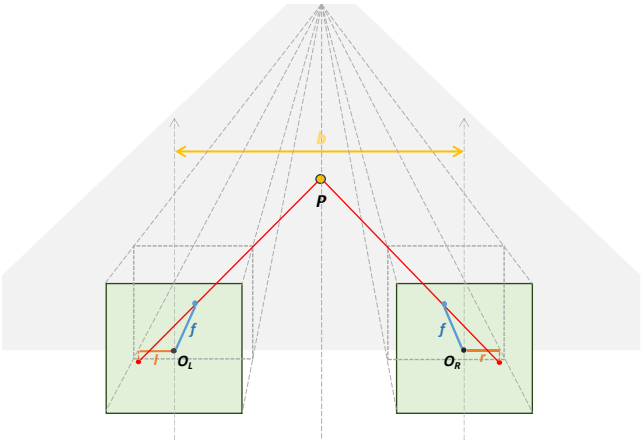

Figure 4. Geometry of epipolar correction for deriving the depth induced from the binocular parallax.

Edge Detection: Edge detection in the primary visual cortex is basically an image processing for finding the object boundaries, therefore allowing us to identify the shapes and recognize the objects. It is more sensitive to the relative difference in color and brightness. To numerically model this function, the three most used edge detection algorithms, *i.e.*, Canny, Prewitt, and Laplacian of Gaussian detection, have been tried and compared. A rule of thumb is that the Canny algorithm outperforms the other two due to its denoising, resistance to edge loss, and ability to preserve object details.

Saliency Detection: The primary visual cortex also generates the saliency maps to highlight the objects, to which our eyes are most likely to be focused. To imitate this function, two static saliency detection algorithms are adopted. The first is based on the visual saliency features, which leverages the integral of the original image to extract high-quality features. The second is based on the spectral residual.

4. Results and Discussion

Camera Settings: Unity is employed to create the three-dimensional scenes to emulate the physical world. As shown in Fig. 5, the left and right images are captured. Two physical cameras are set up to capture perspectives that closely resemble the left and right eyes when directing attention towards specific objects. Settings of interpupillary distance (R/L camera distance), diameter

21680159, 2024, 1, Downloaded from https://sid.onlinelibrary.wiley.com/doi/10.1002/sdtp.17580 by Shanghai Jiaotong University, Wiley Online Library on [03/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the eyeball (focal length), field of view (FOV), projected sensor size, and resolution (number of pixels) are listed in Table 2. These settings mimic the optical parameters of eye [12].

Table 2. Camera settings to generate input images.

Parameter of camera	Parameter of eye	Value
R/L camera distance	Interpupillary distance	60 mm
Focal length	Diameter of eyeball	25 mm
FOV	FOV	150°
Sensor size	Flattened retina size	186 mm
Resolution (pixels)	Number of ganglions	1860 × 1860

Figure 5. Images captured by (a) left and (b) right cameras.

Bilateral Neural Image Fusion: Two types of fusion process are considered: unilateral neural image fusion and bilateral neural image fusion. As shown in Fig. 6, the process of implementing and simulating unilateral image fusion can be divided into four distinct stages. Initially, retinal images are acquired. Epipolar rectification is performed with the aim of aligning the images to a common plane (for the case when sights intersect). In the initialization step, the effect of different positions of the left and right eye is eliminated using epipolar rectification and calibration. Afterwards, the rectified images are split into left or right hemiretinal images, depending on the location of the blind spot. Finally, the hemiretinal images are merged for each hemisphere. The result of unilateral neural image fusion contains information from both sides and is therefore more informative. Bilateral neural image fusion, as shown in Fig. 7, refers to the procedure of merging two unilateral images obtained through unilateral fusion into a unified fused image, containing information from both hemifields. The ocular-dominance-based image fusion is applied to handle the overlap issue during the overlapping between the left and right visual fields.

Depth Map: We shall limit the depth cue to the binocular parallax. Since the binocular parallax are color irrelevant, the input images are monochromatic. As shown in Fig. 8, the left and right images are first rectified using the epipolar correction. The depth in the overlapping visual region can be derived by

$$d = \frac{bfL}{pS} \tag{2}$$

where the parallax p is calculated by the matching algorithm, the projected sensor size S is 186 mm, and the pixel length L is 1860. Other values can be found out in the above camera settings. Firstly, a FOV mask is calculated, which indicates the area of overlapping visual region. Secondly, the calculation of binocular parallax is performed. Two pixels that correspond to the same point in the scene should be matched for the calculation of binocular parallax. The matching algorithm used in our experiment is called SGBM, which is convenient to implement with the OpenCV library. It is worth mentioning that the epipolar correction is a critical step, without which the results will be greatly affected. The detected depth map alongside the ground truth is shown in Fig. 9. The depths of most areas are close to the ground truth. However, in some areas, *e.g.*, the lower right corner of the ground and parts of the wall and sky, the values of depth are missing or inaccurate.

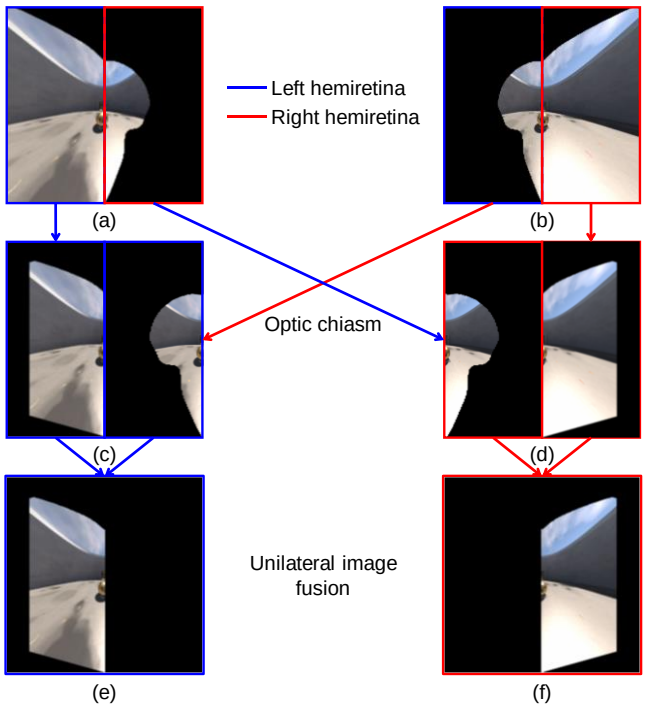

Figure 6. Unilateral neural image fusion: (a) Left retinal image, (b) right retinal image, (c) left hemiretinal images of both eyes, (d) right hemiretinal images of both eyes, (e) left hemisphere image, and (f) right hemisphere image.

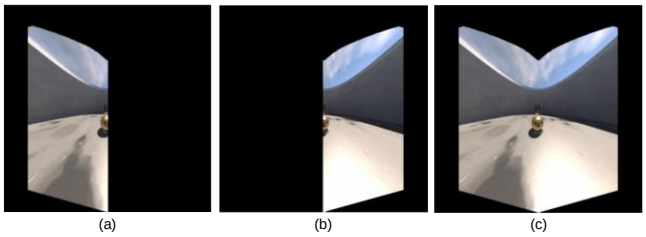

Figure 7. Bilateral neural image fusion: (a) Left hemisphere image, (b) right hemisphere image, and (c) bilateral neural image after fusion.

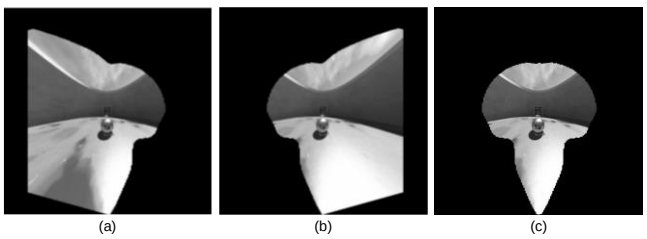

Figure 8. (a) Rectified left eye image, (b) rectified right eye image, and (c) overlap of left and right eye images.

21680159, 2024, 1, Downloaded from https://sid.onlinelibrary.wiley.com/doi/10.1002/sdtp.17580 by Shanghai Jiaotong University, Wiley Online Library on [03/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

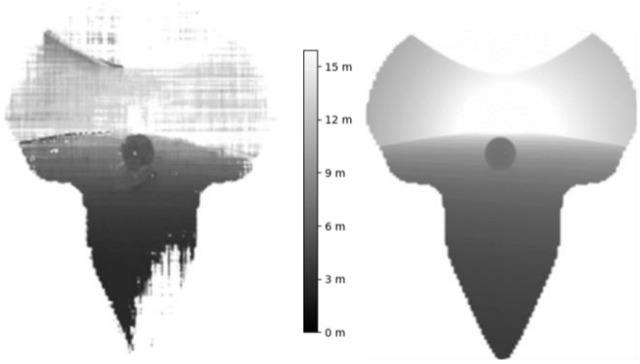

Figure 9. (a) Calculated depth map and (b) ground truth.

Edge Map: Sharpening is an essential preprocessing step for edge detection, as it can enhance the edges of objects. One sharpening technique, called unsharp masking, is adopted. The simulation results of edge detection are shown in Fig. 10. Overall, the edge map with sharpening shows more details than the one without sharpening. For example, the unsharpened detection algorithm ignores most of the clouds in the sky and shadows on the ground, while a sharpened algorithm shows these edges.

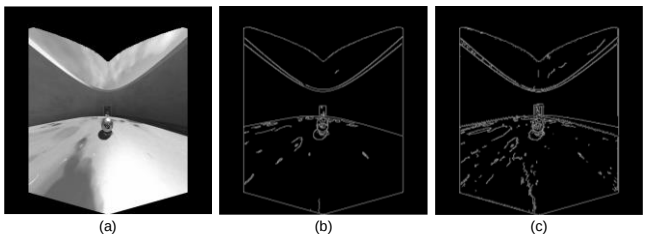

Figure 10. (a) Original bilateral neural image, (b) edge map without sharpening, and (c) edge map with sharpening.

Saliency Map: Both visual saliency feature-based and spectral residual-based saliency detection algorithms are applied to draw the saliency map. For the visual saliency feature-based algorithm, as shown in Fig. 11(a), the bright spot on the ball, shadow of the ball, clouds in the sky, and edges of the scene are marked as salient regions. For the spectral residual-based algorithm, as shown in Fig. 11(b), the ball, and edges of the scene are marked as salient regions. Given the ball is arguably the only one meaningful object in the scene and it sits in the center, it shall be the most salient thing. For this reason, we tend to assume that the latter is more accurate as the shadow of ball in the former does not make much sense. Incidentally, if the motion saliency detection is adopted and the input is a video, the static objects, *e.g.*, the edges of scene, could be excluded.

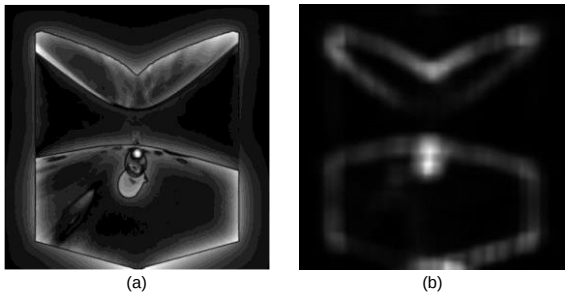

Figure 11. Saliency maps detected by (a) visual saliency feature-based and (b) spectral residual-based algorithms.

5. Conclusions

An artificial retina-based metaverse in conjunction with bionic vision processing has been conceptually probed. To sum up, we would like to condense this work into the following contributions. Contribution 1: herald a next chapter of metaverse hardware, *i.e.*, the retinal implants, for the next-gen humanity. Contribution 2: provide a methodology to simulate the step-wise images along the visual pathway, which are the monocular retinal image, unilateral neural image, bilateral neural image, binocular parallax-induced depth map, edge map, and saliency map. Contribution 3: refine the foveated blurring by exclusively considering midget cells. Admittedly, to supersede the conventional devices, implantable devices still have a long road ahead. Issues like the postoperative reaction, biological incompatibility, and cybersickness [13] remain to be addressed. Nevertheless, we believe this technology will become popularized someday and give us a better metaverse.

6. Acknowledgements

This work is sponsored by National Natural Science Foundation of China (61831015), and Natural Science Foundation of Chongqing (cstb2023nscq-msx0465, cstc2021jcyj-msxmX1136).

7. References

1. L. Zhou, C. P. Chen, Y. Wu, Z. Zhang, K. Wang, B. Yu, Y. Li, "See-through near-eye displays enabling vision correction," Opt. Express **25**(3), 2130–2142 (2017).
2. C. P. Chen, L. Zhou, J. Ge, Y. Wu, L. Mi, Y. Wu, B. Yu, Y. Li, "Design of retinal projection displays enabling vision correction," Opt. Express **25**(23), 28223–28235 (2017).
3. Y. Wu, C. P. Chen, L. Mi, W. Zhang, J. Zhao, Y. Lu, W. Guo, B. Yu, Y. Li, N. Maitlo, "Design of retinal-projection-based near-eye display with contact lens," Opt. Express **26**(9), 11553–11567 (2018).
4. C. P. Chen, L. Mi, W. Zhang, J. Ye, G. Li, "Waveguide-based near-eye display with dual-channel exit pupil expander," Displays **67**, 101998 (2021).
5. C. P. Chen, Y. Cui, Y. Chen, S. Meng, Y. Sun, C. Mao, Q. Chu, "Near-eye display with a triple-channel waveguide for metaverse," Opt. Express **30**(17), 31256–31266 (2022).
6. C. P. Chen, X. Ma, S. P. Zou, T. Liu, Q. Chu, H. Hu, Y. Cui, "Quad-channel waveguide-based near-eye display for metaverse," Displays **81**, 102582 (2023).
7. S. P. Zou, N. Xi, J. Ye, C. P. Chen, Q. Chu, H. Hu, "Intraocular augmented reality display with retinal prosthesis," Proc. SPIE **12443**, 1244306 (2023).
8. J. Jang, H. Kim, Y. M. Song, J.-U. Park, "Implantation of electronic visual prosthesis for blindness restoration," Opt. Mater. Express **9**(10), 3878–3894 (2019).
9. N. Xi, J. Ye, C. P. Chen, Q. Chu, H. Hu, S. P. Zou, "Implantable metaverse with retinal prostheses and bionic vision processing," Opt. Express **31**(2), 1079–1091 (2023).
10. A. B. Watson, "A formula for human retinal ganglion cell receptive field density as a function of visual field location," J. Vision **14**(7), 15 (2014).
11. J. Chen, L. Mi, C. P. Chen, H. Liu, J. Jiang, W. Zhang, "Design of foveated contact lens display for augmented reality," Opt. Express **27**(26), 38204–38219 (2019).
12. L. Mi, C. P. Chen, Y. Lu, W. Zhang, J. Chen, N. Maitlo, "Design of lensless retinal scanning display with diffractive optical element," Opt. Express **27**(15), 20493–20507 (2019).
13. G. Li, M. McGill, S. Brewster, C. P. Chen, J. A. Anguera, A. Gazzaley, F. Pollick, "Multimodal biosensing for vestibular network-based cybersickness detection," IEEE J. Biomed. Health Inform. **26**(6), 2469–2480 (2022).

21680159, 2024, 1, Downloaded from https://sid.onlinelibrary.wiley.com/doi/10.1002/sdtp.17580 by Shanghai Jiaotong University, Wiley Online Library on [03/08/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License