

Training Conditions for Establishing Visual Level 2 Perspective-Taking in Children with Autism Spectrum Disorder

Hiroshi Asaoka^{1*} , Aya Fujiwara², Fumiyuki Noro³

¹Research and Education Faculty, Humanities and Social Science Cluster, Education Unit, Kochi University, Kochi, Japan

²Center for Diversity, Accessibility and Career Development, University of Tsukuba, Ibaraki, Japan

³Faculty of Human Sciences, University of Tsukuba, Ibaraki, Japan

Email: *asaoka@kochi-u.ac.jp

How to cite this paper: Asaoka, H., Fujiwara, A., & Noro, F. (2020). Training Conditions for Establishing Visual Level 2 Perspective-Taking in Children with Autism Spectrum Disorder. *Psychology*, 11, 908-935. <https://doi.org/10.4236/psych.2020.116059>

Received: May 16, 2020

Accepted: June 15, 2020

Published: June 18, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Perspective transformations that are considered the first step of visual level 2 perspective-taking (VPT2) are not highly active in children with autism spectrum disorder (ASD). We investigated training procedures to develop moving to another viewpoint and observing the view from the new viewpoint to establish VPT2. The Face Rotation Task was used as the VPT2 task in three experiments. Experiments 1 and 2 examined the time-lag from the removal of sample stimuli to the presentation of comparison stimuli. The results of the two experiments indicated that VPT2 was established when participants move from the self-viewpoint to the other's viewpoint while continuing to look at the stimuli, and when participants move from the other's viewpoint to the self-viewpoint while maintaining a representation from the other's viewpoint. Eye movements during the presentation of sample stimuli were analyzed in Experiment 3. The results indicated the possibility of mutually looking at elements of a spatial array while performing the perspective transformation.

Keywords

Autism Spectrum Disorder, Visual Level 2 Perspective-Taking, Perspective Transformations, Training, Eye Tracking

1. Introduction

Children with autism spectrum disorder (ASD) tend to imitate another person's behaviors based on the visual appearance from their perspective; for instance, a child will wave his/her hand with his/her palm facing him/her when a mother

waves her hand (e.g., Avikainen, Wohlschläger, Liuhanen, Hänninen, & Hari, 2003; Hobson & Hobson, 2008; Meyer & Hobson, 2004; Ohta, 1987; Smith, 1998). It has been clarified suggested that the factor is that *perspective transformation*, which refers to imaging that where you are in a different location or have a different orientation (Yu & Zacks, 2010), does not fully activate (Conson et al., 2016; Conson et al., 2013) in children with ASD. For example, Asaoka et al. (2019) conducted a task in which children with ASD listened to navigational information and operated a miniature car on a map. The children were instructed to “Turn left, or right,” or “Look from the driver’s viewpoint and turn left, or right” at an intersection. The results showed that children with ASD tended to demonstrate significantly better performance when the viewpoint was explicitly instructed. These findings confirm the inactivation of perspective transformation in children with ASD.

The information used in VPT includes the position of the viewer and the target and the position of objects in the environment in relation to the self and others (Kessler & Thomson, 2010). Perspective transformations are positioned as the first step in the visual level 2 perspective-taking process (VPT2) (Yu & Zacks, 2010), which is the ability to infer how others are looking at an object from a different perspective (Flavell, Everett, Croft, & Flavell, 1981). By transforming ourselves to a different point in space, it becomes possible to judge what is on another person’s left or right, or to make predictions about how things may appear from a different perspective (Pearson, Marsh, Hamilton, & Ropar, 2014). For example, in order for a child to wave his/her hand with the back of hand facing her/him, perspective transformation allows the child to transform his/her perspective to the mother’ perspective. VPT2 is a series of cognitive processes with perspective transformations to imagine the back of the mother’s hand. Incidentally, visual level 1 perspective-taking (VPT1), which develops prior to VPT2, requires only visual information such as the gaze of others to judge whether another person can see an object or not (Flavell et al., 1981), and perspective transformations are not required (Kessler & Rutherford, 2010). Thus, VPT1 is the cognitive process that subserves verbal localizations using “in front” and “behind,” while VPT2 subserves “left” and “right” from a perspective (Kessler & Rutherford, 2010). For instance, Watanabe (2000) conducted the Face Rotation Task in typically developing children, ages 2-years-6-month to 4-years-5-month-old. In task one, the left or right eye, was shown in a face, rotated 90°, 180°, or 270° from the upright position, was highlighted and a buzzer sound was simultaneously presented. The participants were asked to memorize the combination of the highlighted eye and sound. When the face stimulus was set back to the normal upright position, the buzzer sound was presented, and the participants were asked to look or point at the position of the expected highlighted eye.

Previous studies have reported that children with ASD show no difficulty with VPT1, but VPT2 proves especially difficult in perspective transformations (e.g., Baron-Cohen, 1989; Hamilton, Brindley, & Frith, 2009; Hobson, 1984; Leekam,

Baron-Cohen, Perrett, Milders, & Brown, 1997; Leslie & Frith, 1988; Reed & Peterson, 1990; Yirmiya, Sigman, & Zacks, 1994). However, the training conditions for establishing VPT2 have not been sufficiently clear (see Pearson, Ropar, & Hamilton, 2013; for a review). What are the conditions to promote the establishment of VPT2 in children with ASD? In this study, we focused on the cognitive process of VPT2 and hypothesized that the impairment of perspective transformations (Conson et al., 2016; Conson et al., 2013) is due to the difficulty in coding a relation between some elements in the spatial array (e.g., a ball is on his right side) (Huttenlocher & Presson, 1979).

Regarding cognitive processes in VPT2, it has been demonstrated that *physically* turning a body to another viewpoint (Kessler & Thomson, 2010; Kessler & Rutherford, 2010) or moving there (Asaoka, Kumagai, Okamura, & Watanabe, 2016) encourages perspective transformations, that is, *mentally* performing them. Since the mental distance between the self and another viewpoint is short or zero (i.e., the self and another viewpoint coincide), it is easy to mentally simulate the body movement and code the spatial array from another viewpoint (angular disparity effect) (Kessler & Thomson, 2010). However, in the case of physical operation, that is, when moving to another viewpoint, observing a view from there, returning to the self-viewpoint, and recalling the view, a longer time-lag (hereafter, TL) occurs between the observation and the recollection of the view compared with the mental operation. The TL was not sufficiently considered as a factor influencing the establishment of VPT2 in previous studies in children with ASD (Okuyama & Isawa, 2010). In the meantime, it is necessary to retain spatial memory across viewpoint movements, and the time that can be retained depends on age (Montefinese, Sulpizio, Galati, & Committeri, 2015). Paradoxically, visual representation from another viewpoint is retained by setting the TL according to the cases. Therefore, we expect that the reduction of retention affects the understanding that it becomes easier to infer the spatial array by physically moving the viewpoint. In this study, the *establishment* of VPT2 was defined as the stable occurrence of responses based on the array from another viewpoint, regardless of the physical or mental means of viewpoint movement.

Then, how is the coding of spatial arrays performed when the task setting is introduced that facilitates the activation of perspective transformations? When coding, as shown in “the milk is to the right of the cereal box,” the cereal box is an external landmark required to grasp the spatial position of the milk (Huttenlocher & Presson, 1979). At this time, the spatial array is recoded depending on where the viewpoint is placed. If it assumes that “the milk is to the *right* of the cereal box from the self-viewpoint,” the array is recoded that “the milk is the *left* of the cereal box from the face-to-face perspective.” Furthermore, visual appearance continuously changes in the process of moving the viewpoint; for example, “First, I look at the milk to the *right* of the cereal box. Then, I look to the *side* of the milk or the cereal box. Finally, I look at the milk to the *left* of the cereal box.” Looking at objects from various viewpoints in this way promotes

learning of physical properties between space and objects (Ayres, 1979) and the acquisition of vocabulary (Slone, Smith, & Yu, 2019).

To summarize, we predict that the setting of TL according to the cases and the formation of gaze behaviors to objects during viewpoint movement impacts the establishment of VPT2. In this study, we conducted the Face Rotation Task (Watanabe, 2000) to examine how the conditions under which selective responses from another viewpoint were promoted. The present study consisted of three experiments. In Experiment 1, we investigated the factors of TL that occur from the removal of the sample stimuli to the presentation of the comparison stimuli. Based on the results, in Experiment 2, the effects of the setting of appropriate TL and formation of gaze behaviors to the sample stimuli were examined. In Experiment 3, eye movements at the presentation of the sample stimuli were analyzed to verify the effectiveness of the conditions suggested by Experiment 1 and 2.

2. Experiment 1

We examined whether reducing the TL from an observation of another viewpoint to the presentation of the comparison stimuli would promote the establishment of VPT2.

2.1. Method

2.1.1. Participants

Three Japanese children with ASD (two boys and one girl) participated. The participants were recruited from local and the university clinic centers based on the following three criteria: 1) the child's chronological age (CA) was 3 years and 6 months or more based on the result of Watanabe (2000); 2) an intelligence quotient (IQ) of 70 and over as assessed using the Tanaka Binet Intelligence Scale-Fifth Edition (Tanaka Binet-V; Tanaka Institute for Educational Research, 2003), and 3) the child was diagnosed with ASD by at least one doctor using the standard and diagnostic criteria of the DSM-5 (American Psychiatric Association, 2013) and had a score of ≥ 9 in early childhood or ≥ 13 in childhood on the Parent-interview ASD Rating Scale-Text Revision (PARS-TR; PARS Committee, 2013). The sub and total scores of PARS has correlations with the domain and total scores of the Autism Diagnostic Interview-Revised (ADI-R; Lord, Rutter, & Le Couteur, 1994), indicating the convergent validity of PARS (Ito et al., 2012). The PARS-TR peak symptoms scale comprises 34 items in early childhood and 33 items in childhood that describe the characteristic behavioral symptoms of ASD (e.g., "Has the child ever had difficulty making eye contact?"), with a scale of 0 to 2 for each item.

Yuki was a 4-year-2-month-old girl diagnosed with ASD who was enrolled in a regular kindergarten. On the Tanaka Binet, Yuki's IQ was 96, her mental age (MA) was 3:8, tested at CA 3:10. Her PARS-TR peak symptoms scaled score was 28 points. Ken was a 6-year-0-month old boy diagnosed with ASD and ADHD

who was also enrolled in a regular kindergarten. On the Tanaka Binet, Ken's IQ was 97, his MA was 5:6, tested at CA 5:8. His PARS-TR peak symptoms scaled score was 18 points. Taro was a 6-year-1-month old boy diagnosed with ASD who was enrolled in special needs classes in an elementary school. On the Tanaka Binet, Taro's IQ was 81 his MA was 4:2, tested at CA 5:2. His PARS-TR peak symptoms scaled score was 13 points. The parents of each participant provided informed consent form for this study, which was approved by the research ethics committee of the Faculty of Human Sciences, University of Tsukuba (No. 28-223). All participants were compensated for their time.

2.1.2. Setting and Materials

All conditions (baseline; BL, pre-training, large/middle/small TL, probe, and maintenance) were conducted in the therapy room of the university clinical center. Except for the large TL condition, an experimental stimulus was used with the participant and the experimenter who sat on the carpet facing each other. In the large TL condition, the experimental stimulus was placed on a desk, and they were seated facing each other. At this time, the experimenter sat in a slightly separated position so as not to disturb the movement of the participant. Two video cameras with wide-angle lenses were set in the room and participants were shot during the experiment.

The experimental stimuli were a symmetrical facial model or image (**Figure 1**). The facial model was used in the BL, pre-training, large/small TL, and maintenance phases. The facial image was used in the small TL and probe. The facial figures were used as stimuli because of their clear vertical direction and familiarity to children of the target age (Watanabe, 2000). The facial model was a wooden disk with a diameter of 40 cm and a thickness of 2.6 cm, and two holes with a diameter of 5 cm were drilled at a target position 10 cm apart from the center. Two 3.3 cm diameter lights that can be turned on and off by switch were embedded in them. The light-emitting part was pale white and shone with glitter. The nose, mouth, and hair were made by cutting out felt and pasting it on the disk. In addition, the facial image was approximately 1/2 of this model, projected on the screen of a 21.5-inch personal computer (Sony SVT2121A1J) set facing upward on the floor. The computer and a small remote controller (Kokuyo ELA-FP1) were connected wirelessly, and the image could be changed by operating the switch at the hands of the experimenter.

2.1.3. Experimental Design

The experimental design was a multiple baseline across participants design (Ledford & Gast, 2018).

2.1.4. Procedure

Baseline. The angle of the part corresponding to the jaw of the facial stimulus closest to the participant is expressed as 0°, and the positions rotated counter-clockwise by 90° are expressed as 90°, 180°, and 270° (**Figure 1**). Initially, the experimenter turned on either the light of right or left eye of the facial stimuli

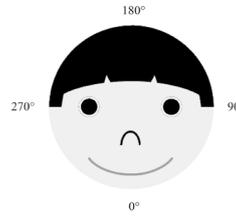


Figure 1. Face rotation task stimulus.

(hereafter referred to as the sample stimuli) and presented it randomly. After the participant observed the sample stimuli, the experimenter removed it, turned the light off, and presented the facial stimuli in a state where it turned to 0° (hereafter, the comparison stimuli). At this time, the participant's eyes were closed to prevent his/her from seeing the rotation of the stimulus. Immediately, the experimenter asked, "Which eye was shining?" and the participant pointed or touched either eye. This question was omitted when the participant understood the task. The experimenter proceeded to the next trial with or without a correct response. The presentation angle of the sample stimuli (hereafter, rotation angle) and the lighting direction of the eye were counterbalanced between trials. A total of 16 trials in each block were conducted. Each block consisted of 4 trials randomly arranged by rotation angles of 0° , 90° , 180° and 270° . Each rotation angle included 2 trials by the shining of left/right eye.

Pre-training. It was conducted for Ken because his average number of correct responses at a rotation angle of 0° was below the chance level in the whole BL. The experimenter presented the sample stimuli at a rotation angle of 0° in all trials. After Ken looked at the sample stimuli, the experimenter presented the comparison stimuli, and Ken made a choice. When a correct response occurred, the experimenter provided verbal praise, clapping hands, or used physical interaction, such as shaking his/her hands. When an error response occurred, the experimenter pointed to the correct eye and said, "The eye was shining." Each block consisted of 4 trials with a rotation angle of 0° and 2 trials by the shining of left/right eye.

Time-lag conditions. The experimenter presented the sample stimuli at a rotation angle of 90° or 270° . When the error response occurred, the experimenter presented it again in the same angle. Then, the participant moved from the original position to a position where the facial stimuli appeared to be upright (hereafter, another viewpoint) and observed it from there (Figure 2). The experimenter checked whether the participant was looking at it when the participant was at another viewpoint, and if the participant's face was not directed at it, the experimenter provided a verbal (e.g., "Look at the face of the model") and/or physical prompt. However, no prompt was presented while moving. Then, the participant returned to the original position, and the experimenter presented the comparison stimuli. When the correct response occurred, the experimenter provided verbal praise, clapped hands, or used physical interaction. When the error response occurred, the experimenter pointed to the correct eye. Other procedures were the same as those of the BL.

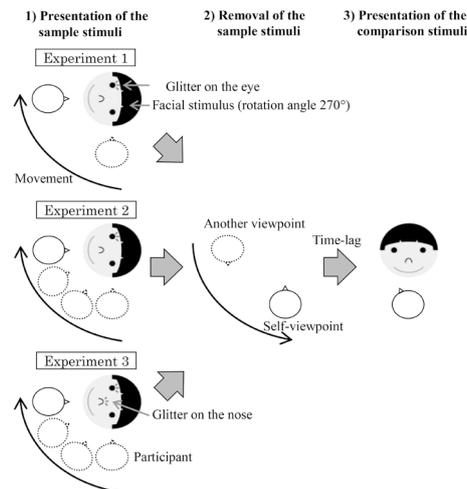


Figure 2. Corrective procedures in Experiment 1, 2 and mutual gaze condition in Experiment 3.

The TL was defined as the time interval from the time the participant started moving with the return to the original position until the time the experimenter presented the comparison stimuli. The following three conditions were set as the corrective procedures so that the TL gradually decreased. First, in *the large TL condition* (hereafter, *LTL*), the participant moved on foot. Second, in *the middle TL condition* (hereafter, *MTL*), the movement method was changed from the LTL. The participant turned his/her face and body to another viewpoint while sitting on the carpet. If the participant's face and body were not sufficiently directed to the sample stimuli, the experimenter provided a model and/or physical prompt. Third, in *the Small TL condition* (hereafter, *STL*), the method of presenting the stimuli was changed from the MTL phase. The facial images were presented and removed immediately by using the PC. The experimenter presented the sample and comparison stimuli on the screen by using presentation software (Microsoft PowerPoint 2016). A black slide was inserted between the removal of the sample stimuli and the presentation of the comparison stimuli. In this condition, the facial stimuli were not rotated and the lights were not turned on/off so that the participant did not close their eyes. A total of eight trials in each block were conducted. Each block consisted of 4 trials randomly arranged by rotation angles of 90° and 270°. Each rotation angle included two trials by the shining of left/right eye. In principle, the transition criteria for the conditions were 15 trials (62.5%) or less of 24 trials (12 trials each presenting sample stimuli at 90° and 270°) in 3 consecutive blocks. The criteria of accomplishment were the occurrence of correct responses in 22 of 24 trials (91.6%) in 3 consecutive blocks. Furthermore, the STL without feedback (FB) condition was conducted only for Yuki. The procedures were the same as those of the STL, except that the experimenter presented the sample stimuli at four rotation angles and did not provide performance feedback.

Probe. The procedures of the probe were the same as that of the BL phase.

Maintenance. The maintenance phase was conducted for Ken and Taro. One

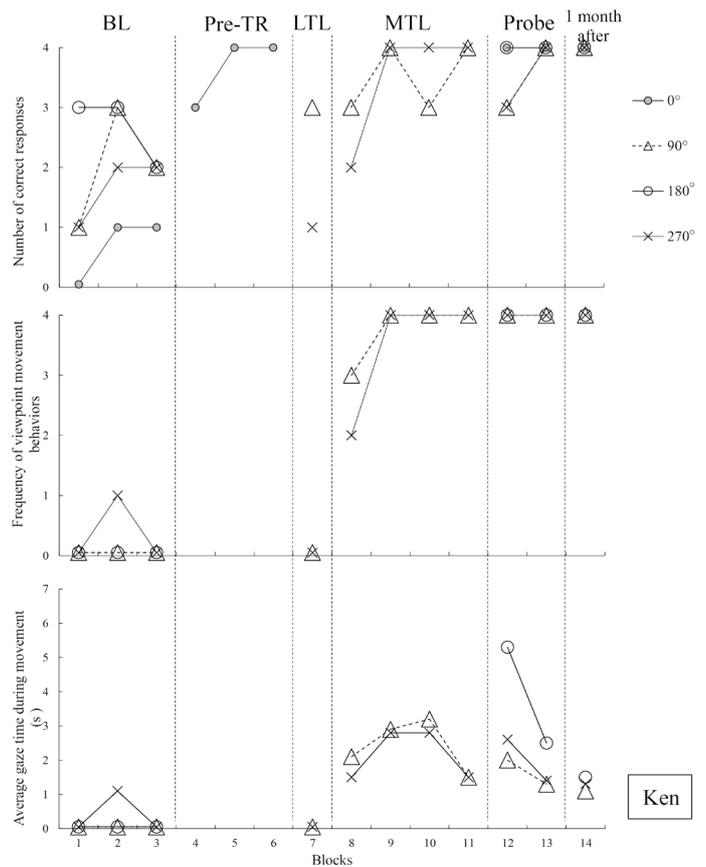
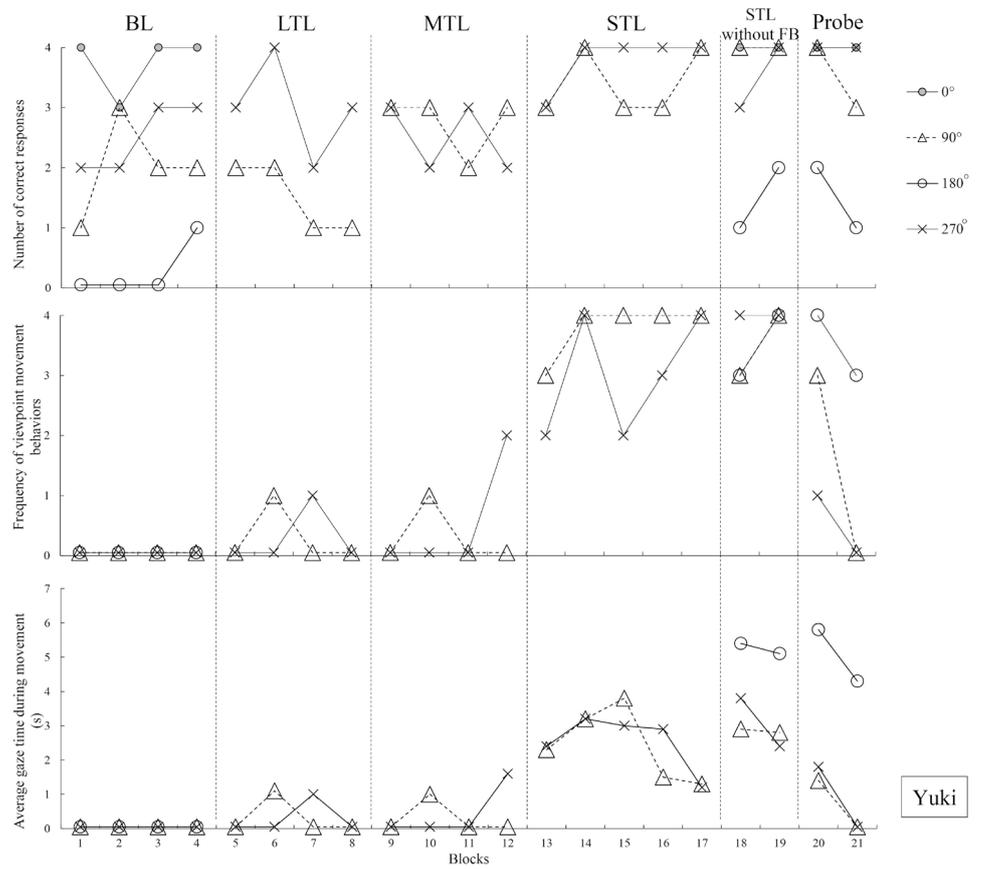
month after completion of the probe, the experimenter confirmed whether the training effects were maintained in the same procedure as in the BL phase. Maintenance was not conducted for Yuki because her performance decreased in the probe.

2.1.5. Dependent Variables and Data Analysis

Dependent variables were classified and defined into the following four categories, each of which was quantitatively analyzed using video editing software (Adobe After Effects CC). 1) *Number of correct responses*: The number of times the participant chose the shining eye was counted for each rotation angle. 2) *Frequency of viewpoint movement behaviors*: A line segment was placed on the video to bisect the participant's torso from the head and the video data at the time of presentation of the sample and comparison stimuli was played back frame by frame every 0.1 s. In the analysis, the viewpoint movement behavior was defined as follows. First, using the line segment as a guide, the participant moved his/her face and body in the direction of another viewpoint when the sample/comparison stimuli were presented. Second, the participant's face was directed at the facial stimuli. Third, the sum of these times is 1 s or more. When all three conditions were satisfied, it was assumed that the viewpoint movement behavior has occurred, and the frequency was calculated for each rotation angle. 3) *Average gaze time during movement (s)*: The time to gaze the stimuli was calculated by subtracting the time moving toward another viewpoint from the time immediately before moving to return to the original position. The average gaze time during movement was calculated by dividing the total gaze time of 1 block by the frequency of viewpoint movement behaviors. 4) *Average time-lag in each TL condition (s)*: The TL was calculated by subtracting the time to the return of the original position from the time when the comparison stimuli were presented for each trial when the corrective procedures were introduced. These sums were calculated by dividing by the number of error responses across each TL condition.

2.2. Results

Yuki. The number of correct responses in the BL phase was highest when the sample stimuli were presented at 0° (hereafter simply referred to as 0°), followed by 270°, 90°, and 180° (**Figure 3**). The viewpoint movement behavior did not occur. The average TL in the LTL and MTL were 7.1 s (ranging from 6.0 - 8.7 s) and 3.0 s (2.7 - 3.9 s), respectively. There were no positive changes in the number of correct responses, frequency of viewpoint movement behaviors, and average gaze time during movement at 90° and 270° in these conditions. A behavioral feature of note was that she often looked at people and things around her while moving to another viewpoint. Furthermore, at the time of the retrial in the MTL phase, she often leaned her body left/right while looking at the experimenter, and then pulled her chin to see the stimuli while holding the posture. The average TL in the STL was 1.0 s (0.9 - 1.0 s), with an increase in the number of



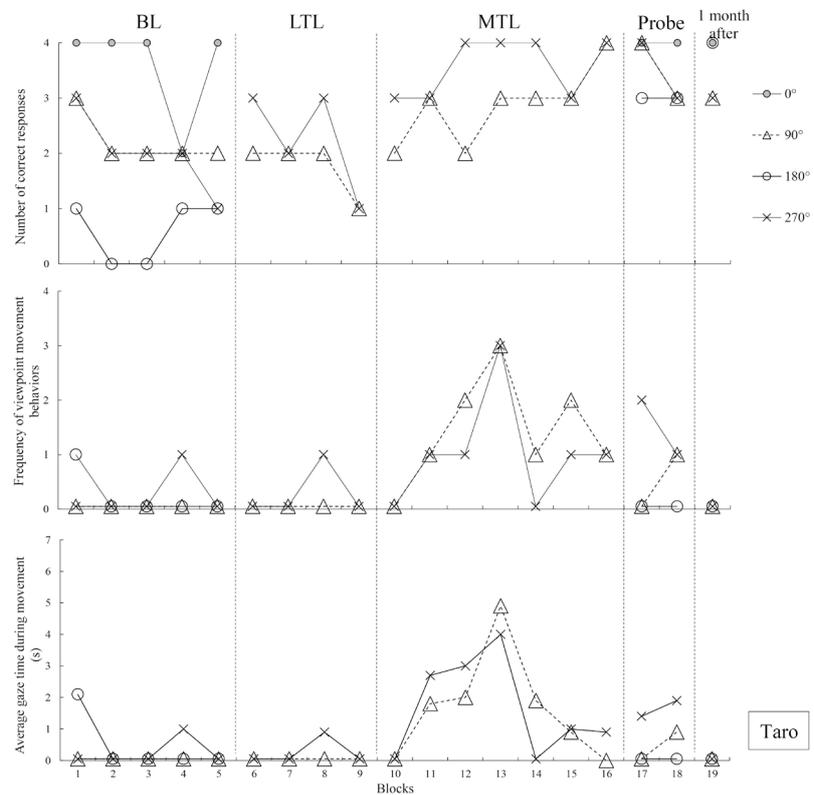


Figure 3. Results of the Face Rotation Task in Experiment 1. Note: BL; Baseline, Pre-TR; Pre-training, LTL; Long time-lag, MTL; Middle time-lag, STL; Short time-lag.

correct responses, and frequency of viewpoint movement behaviors at 90° and 270°. Additionally, with an increase in these responses, the average gaze time increased and then tended to decrease. In the STL phase without FB and probe, the number of correct responses and frequency of viewpoint movement behaviors at 90° and 270° increased compared to the BL. However, the number of correct responses increased only slightly at 180° under these conditions (1 - 2 times). She often expressed “I forgot.” when choosing the comparison stimuli at 180°.

Ken. The number of correct responses in the BL phase was high at all positions, 180°, 90°, 270°, and 0°. The viewpoint movement behavior occurred once at 270° throughout the condition. In pre-training, he fulfilled the criteria of accomplishment in 3 blocks. The average TL in the LTL was 5.5 s. When performing a retrieval, he cried loudly and bit his own hands. Although he fulfilled the transition criteria, we judged that it would be a great burden for him to continue the condition, and we stopped after 1 block. The average TL in the MTL was 3.1 s (2.1 - 3.4 s). Similar to the results of Yuki’s STL, the correct responses and viewpoint movement behaviors at 90° and 270° continued to occur stably, and average gaze time showed a bell-shaped curve tendency. With regard to the viewpoint movement behavior, he put his hands on his left/right front and then stretched out his arms and back while continuing to look at the sample stimuli. In the probe and maintenance phases, the number of correct responses and viewpoint movement behaviors occurred in most trials. At this time, the average

gaze time tended to decrease as well as the MTL.

Taro. The number of correct responses in the BL phase was as high at 0°, 90°, 270°, and 180° in this order. The viewpoint movement occurred once each at 180° and 270° in the whole condition. The average TL in the LTL was 4.5 s (3.9 - 5.9 s). As with Yuki and Ken, there were no positive changes in the dependent variables under this condition. The average TL in the MTL was 3.0 s (2.0 - 6.1 s), and the number of correct responses, frequency of viewpoint movement behaviors, and average gaze time at 90° and 270° increased, and then the frequency of viewpoint movement behaviors and average gaze time decreased. He moved the viewpoint in the same way Ken did. In the probe and maintenance phases, the number of correct responses increased in comparison with the BL. At this time, viewpoint movement behaviors occurred stably in Ken, whereas they occurred less frequently in Taro.

2.3. Discussion

We examined whether changing the TL for retaining visual representations would promote the establishment of VPT2. The TL conditions including the number of correct responses, frequency of viewpoint movement behaviors, and average gaze time impacted the establishment of VPT2. The results support the suggestion that the TL affects VPT2's performance (Okuyama & Isawa, 2010). Yuki's CA was about 2 years lower than that of Ken and Taro, and the time for retaining representations can be lower depending on the CA (Montefinese et al., 2015). This is supported by the positive changes of the dependent variables in Yuki's STL compared to LTL and MTL, and her expression such as "I forgot" in the STL without FB and probe. Since the center of VPT2 is perspective transformations, it is necessary to assess the TL before beginning the intervention and to introduce the TL conditions for each case.

An increase in the average gaze time supports an indication of the importance of experiencing changes in visual appearance associated with body movement (Ayles, 1979; Slone et al., 2019). In Experiment 1, we examined only whether the participant was looking at the sample stimuli from another viewpoint and did not prompt while moving to another viewpoint. Why did the gaze time increase despite this? One possibility is that the movement method of the MTL and STL is affected. Since the participants are in a fixed sitting position in this method, it can be said that it is easier to pay attention to the stimuli than to move to another viewpoint. This is suggested by Ken and Taro's viewpoint movement behaviors, in which they moved their face and body in the direction of another viewpoint while looking the stimuli. Moreover, we assume that the presentation of verbal praise and clamping hands in association with the occurrence of viewpoint movement behaviors and correct responses functioned as reinforcing stimuli and increased their gaze time. The decline after the increase suggests that the participants may have learned to code the relation between components (Huttenlocher & Presson, 1979)—that is, the spatial array between the glittering eye and other parts (hair, nose and/or mouth). The reduction is considered to be a

step leading towards considering another viewpoint and inferring the appearance from there. Therefore, future studies should examine the effect of moving to another viewpoint while continuing to look at sample stimuli.

Regarding other factors affecting the establishment of VPT2, we consider that the performance of the BL affected the probe. In particular, Yuki's correct response at 180° was only 1 trial. In other words, it is inferred that Yuki had strong spatial egocentrism (Morss, 1987). In addition to memory retention, its strength may have kept correct responses at 180° in her probe. It also suggests an effect of intellectual development. Taro showed that the correct responses and viewpoint movement behaviors did not occur stably in the MTL phase, and the frequency of viewpoint movement behaviors was low in the probe and maintenance phases. Taro's IQ is about 15 lower than that of Yuki and Ken, and Taro may not be able to fully understand the relation that moving the viewpoint makes it easier to infer the appearance from another viewpoint.

In summary, it appears that there is a mutual relationship between the retention of the representation and the viewpoint movement/transformation.

3. Experiment 2

We assessed the time it took for participants to hold the visual representation and introduced TL conditions. At the same time, we examined whether VPT2 could be established with a smaller number of blocks by forming a behavior by moving to another viewpoint while continuing to look the sample stimuli.

3.1. Method

3.1.1. Participants

Two Japanese children with ASD (one boy and girl) participated in Experiment 2. The recruitment methods and selection conditions for the participants were the same as in Experiment 1. Fuku was a 5-year-1-month-old boy diagnosed with ASD who was enrolled in a regular kindergarten. On the Tanaka Binet-V, Fuku's IQ was 100, his MA was 4:8, tested at CA 4:8. His PARS-TR peak symptoms scaled score was 34 points. Moe was a 6-year-2-month girl diagnosed with ASD who was also enrolled in a regular kindergarten. On the Tanaka Binet-V, Moe's IQ was 104, her MA was 6:0 tested at CA 5:9. Her PARS-TR peak symptoms scaled score was 25 points. As with Experiment 1, Experiment 2 was conducted with the approval of the ethics committee and parental consent.

3.1.2. Setting and Materials

In the TL7s condition of the assessment, the PC was placed on the desk and the participants stood in front of it. In the TL 3.5 s and 1 s conditions, the PC was placed on the carpet and the participants sat in front of it. In the BL phase, time-lag and gaze conditions, and probe of Fuku, the PC with the facial image was placed, and he and the experimenter sat on the carpet facing each other. For Moe, the facial model was placed on a desk, and she and the experimenter were seated facing each other. The other settings were the same as in Experiment 1. In

all conditions except the assessment, the facial image and model were used for Fuku and Moe, respectively. The facial image and model, PC, and desk were the same as those used in Experiment 1.

3.1.3. Experimental Design

The experimental design was a multiple baseline across participants design.

3.1.4. Procedures

Assessment for selecting time-lag conditions. Based on the result of the average TL in each TL condition of Experiment 1, three conditions of the TL 7 s, TL 3.5 s, and TL 1 s were set. The TL 7 s was set based on the result that the maximum of the average TL in the LTL for Yuki, Ken, and Taro was 7.1 s. The TL 1 s was set based on the result that the average TL in the STL for Yuki was 1 s. In TL 3.5 s, the average TL in the MTL for three participants was about 3 s, but in order to clarify the time difference between the conditions, the time was set to half the time of the TL 7 s. Initially, all the sample stimuli were displayed on the PC screen at a rotation angle of 0°. Immediately after that, in the TL 7 s condition, the participant walked back 2 to 3 meters laterally in either the left or right direction instructed by the experimenter to correspond to the LTL. After 7 s, the experimenter presented the comparison stimuli and the participant chose. In the TL 3.5 s and 1 s conditions, the experimenter made the participants turn his/her body to the left/right direction to correspond to the MTL/STL and presented the comparison stimuli after 3.5 s or 1 s. A total of 8 trials in each condition were conducted. Each condition consisted of 4 trials by the shining of left/right eye. TL 7 s, 3.5 s, and 1 s conditions were conducted in this order, and the assessment was completed when correct responses occurred in 7 or more trials (88%) in the 1 condition. As a result of the assessment, we selected the STL for Fuku and the LTL for Moe.

Baseline. The procedures were the same as in Experiment 1.

Time-lag and gaze conditions. Conditions were the same as in Experiment 1 except that the TL conditions corresponding to the results of assessment were introduced immediately after the BL and the participants moved to another viewpoint while continuing to look the facial stimuli in the retrieval. The conditions introduced to Fuku and Moe were named *the LTL + gaze* and *STL + gaze conditions*, respectively. The following are the differences from Experiment 1. When the error response occurred, the experimenter instructed Fuku to “Turn your face and body to here while looking the facial image” and Moe to “Walk to here while looking the facial model” Then, the experimenter presented a model who turned the face and body in the direction of another viewpoint, while looking at the facial image of the sitting position for Fuku and a model that walked while looking at the facial model for Moe. Subsequently, he/she imitated it (**Figure 2**). While moving to another viewpoint, the experimenter provided the verbal and/or physical prompt if his/her face and body were not sufficiently directed to the facial stimuli.

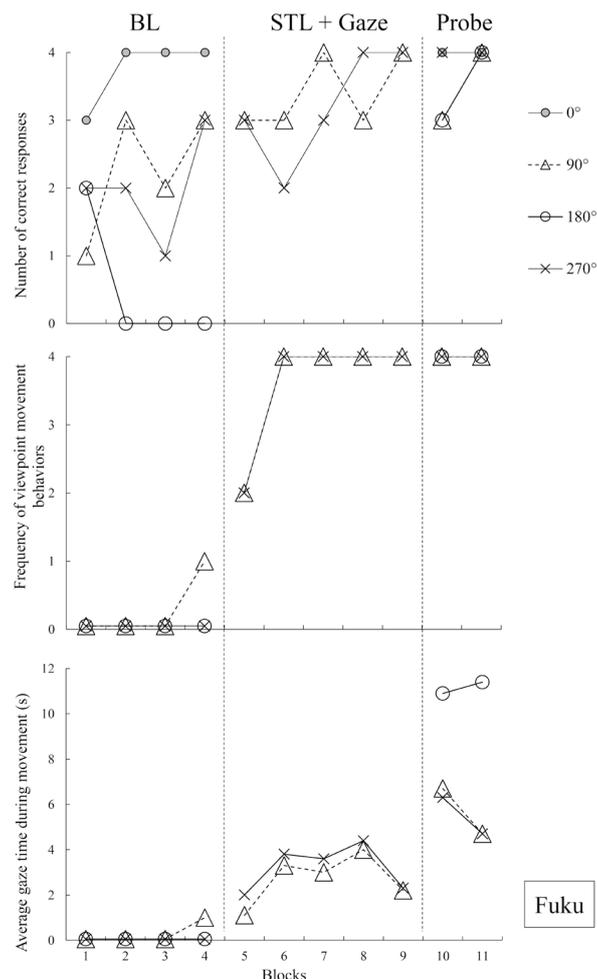
Probe. The procedures of the probe were the same as that of the BL.

3.1.5. Dependent Variables and Data Analysis

The dependent variables and methods of data analysis were the same as in Experiment 1.

3.2. Results

Fuku. In the assessment, the number of correct responses for TL 7 s, 3.5 s, and 1 s were 4, 6, and 8 times, respectively. In the BL, the number of correct responses was high at 0°, 90°, 270°, and 180° in this order (**Figure 4**). The viewpoint movement behavior occurred once at 90° throughout the condition. The average TL in the LSL + gaze was 1.2 s (1.0 - 1.8 s). Similar to the results of three participants in Experiment 1, the correct responses and viewpoint movement behaviors at 90° and 270° continued to occur stably, and average gaze time showed a bell-shaped curve tendency. After the second blocks of this condition, the viewpoint movement behaviors occurred in all trials. Regarding this behavior, he placed his hand diagonally forward while sitting, turned his face and body in the direction of another viewpoint, and looking the sample stimuli. He fulfilled the criteria in 5 blocks. In the probe, the viewpoint movement behaviors occurred in all trials, and the number of correct responses increased at all angles compared to the BL.



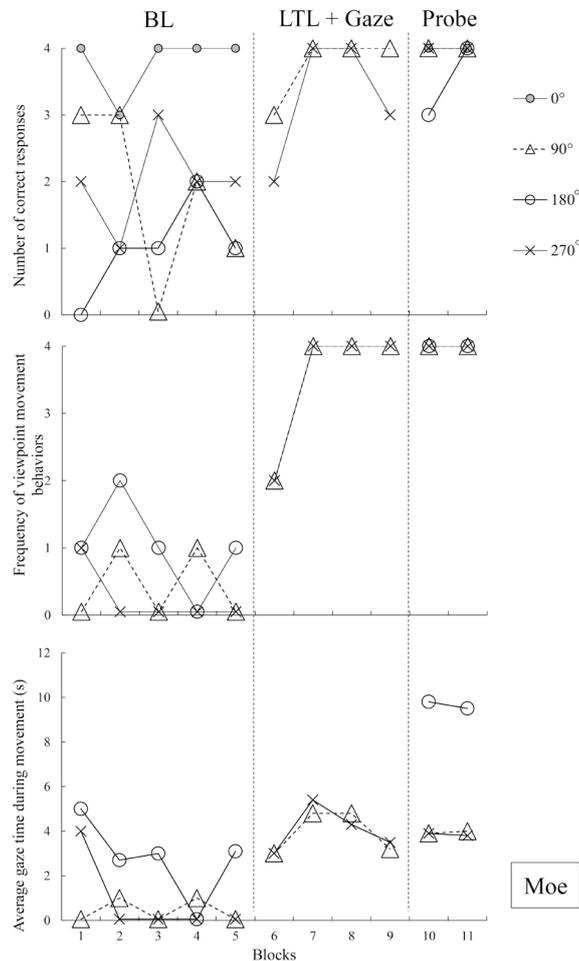


Figure 4. Results of the Face Rotation Task in Experiment 2. Note: BL; Base-line, STL + Gaze; Short time-lag + Gaze, LTL + Gaze; Long time-lag + Gaze.

Moe. In the assessment, the number of correct responses for TL 7 s was 7 times. In the BL phase, the number of correct responses was high at 0°, 270°, 90°, and 180° in this order. The viewpoint movement behaviors occurred in the range of 0 to 2 times for each block. Most of these behaviors occurred when presenting the comparison stimuli. She rotated her body and face in the direction of another viewpoint while sitting. The average TL in the LSL + gaze was 6.2 s (6.1 - 6.2 s). In this condition, the occurrence of correct responses and viewpoint movement behaviors and gaze time showed the same tendency as Fuku. Her viewpoint movement behaviors were to walk to another viewpoint while continuing to look at the presentation of the sample stimuli. She fulfilled the criteria in 4 blocks. The error response occurred in 1 trial at 180° in the entire probe. The viewpoint movement behaviors similar to the LTL + gaze occurred in all trials.

3.3. Discussion

In Experiment 2, we introduced the TL conditions and shaped gaze behaviors to the sample stimuli. As a result, both of them achieved the criteria of accomplishment with a smaller number of blocks than Experiment 1. Furthermore, the

performance of the probe was improved compared to the BL, including that of 180°. With the introduction of STL/LTL + gaze, the number of correct responses increased, and the viewpoint movement behaviors reached 100% in 2 blocks of the TL + gaze conditions in all participants. These results suggest the validity of the setting of TL based on the results of assessment and the effectiveness of gaze behavior's formation. From the result of the LTL + gaze in Moe, it is possible that the occurrence of appropriate selective responses may be promoted, even in the case of the method of moving by walking, if the participants move while observing the stimuli and recollect spatial memory. Therefore, the conditions to promote selective responses based on the appearance from another viewpoint suggest that: 1) the participants move from the self to another viewpoint while continuing to look the stimuli and 2) they move from another to self-viewpoint while retaining the representation from another viewpoint.

The shift from increased to decreased gaze time may be evidence of a transition from a physically to mentally moving viewpoint. In the increase, the participants chronologically observe the appearance that changes from time to time. In other words, they “decouple” a perspective (Hamilton et al., 2009) and observe the facial stimuli from a seamless viewpoint. The results of gaze time supported Ayres (1979) and Slone et al. (2019) findings, as with Experiment 1. Subsequently, gaze time decreases and converges to 0s, indicating that correct responses occur only by mentally putting a self-perspective into another perspective (Takano, 1998). Then, how do participants look at the sample stimuli when observing them in a time series or when inferring the appearance of another viewpoint from self-viewpoint? The Face Rotation Task requires them to code the relation between the eyes and other parts of the facial figure (Watanabe, 2000). From the results of the increase and decrease in gaze time in Experiments 1 and 2, the increase may indicate that the participants repeatedly looked at the entire face, specifically the eyes, and the decrease may indicate that the frequency and time of mutual gaze are decreasing. However, it may indicate that the participants simply memorize a pattern of correct responses according to the presentation angle of the sample stimuli in these processes. An analysis of eye movements is essential to demonstrate that conditions of 1) and 2) promote not only conditional discrimination (Falla & Alós, 2016; Okuyama & Isawa, 2010), but also perspective transformations.

4. Experiment 3

The results of Experiments 1 and 2 suggest that the participants may look at the entire facial figure when performing the Face Rotation Task. Thus, we analyzed the eye movements during the presentation of the sample stimuli and verified the effectiveness of the two conditions in each experiment.

4.1. Method

4.1.1. Participants

Table 1 shows the profiles of the participants. Five participants in Experiments 1

and 2 participated as a follow-up (FU) group. In the ASD group, children with ASD who met the following criteria were newly recruited: 1) the CA is 8 years old. The reason for this was that the children after the age of 8 years are able to perform VPT2 skills spontaneously (Asaoka et al., 2019) and their fluency increased (Elekes, Varga, & Király, 2017), suggesting that the introduction of conditions was expected to be effective; 2) intellectual development is more advanced. All children were assessed for intellectual development with the Picture Vocabulary Test-Revised (PVT-R) (Ueno, Nadeo, & Iinaga, 2008), which is the official standardized scale in Japan. For inclusion, the participants had a scaled PVT-R score of 6 points or above (i.e., above average verbal intelligence); 3) the children had been diagnosed with ASD by at least one doctor; 4) the children could engage in the task wearing wearable eye tracking glasses. Participants Sora, Ryo and Hina met these criteria. The criteria for the typically developing (TD) group were as follows: 1) the CA was 3 years and 6 months or more, and the PVT-R score was 6 points or above; 2) the PARS-TR peak symptoms scaled scores are below the cutoff value; 3) and the conditions for the eye tracker were similar to those of the ASD group. According to these criteria, Jun, Gaku and Anna participated. As with Experiments 1 and 2, Experiment 3 was conducted with the approval of the ethics committee and parental consent.

Table 1. The profiles of the participants in Experiments 3.

Group	Participant	Gender	CA	PVT-R		PARS-TR	
				VA	SS	Early childhood	Childhood
FU	Yuki	F	5; 10	4; 7	10	28	—
	Ken	M	8; 0	7; 5	12	18	24
	Taro	M	7; 6	6; 1	7	13	20
	Fuku	M	6; 0	5; 6	10	34	—
	Moe	F	7; 2	6; 5	9	25	31
Ave.			6; 11	6; 0	9.6	23.6	25.0
ASD	Sora	M	8; 1	8; 9	12	35	19
	Ryo	M	8; 1	10; 2	16	29	20
	Hina	F	8; 3	8; 3	11	11	26
Ave.			8; 2	9; 1	13	25.0	21.7
TD	Jun	M	5; 5	5; 9	10	6	—
	Gaku	M	6; 10	8; 1	14	4	—
	Anna	F	6; 7	6; 5	10	0	0
Ave.			6; 3	6; 9	11.33	3.3	0.0

Note: CA, chronological age; VA, vocabulary age; SS, scaled score; FU, follow-up; ASD, autism spectrum disorder; TD, typically developing; male, M; female, F; Age values are in years; months. The PARS-TR was conducted after the end of the study in Sora. A shortened version of PARS-TR was implemented in Anna.

4.1.2. Setting and Materials

In all conditions, the participants stood in front of a square monitor (Eizo EV2730Q) placed on a table with the screen facing the ceiling. The facial stimulus (**Figure 1**) displayed on the monitor was 40 cm in diameter and the eyes were 5 cm in diameter, which were 10 cm away from the center. Under the table, a PC for the stimulus presentation (NEC PC-LZ650TSS) and a mobile battery (Anker AK-A1701511) for supplying power to the monitor, were present. There was enough space around the monitor for the participants to move. The experimenter stood in front of the PC (Dell Precision M6600) to monitor eye movements. The PC for monitoring was placed so that the screens could not be seen by the participants. It was connected to the wearable eye tracking glasses (Tobii technology Tobii Pro Glasses 2; hereafter, the eye tracker) via Wi-Fi. The monitoring PC displayed the camera image of the eye tracker and the gazing point. The video and its position were the same as those in Experiments 1 and 2. The reasons why the participants stood and walked were that the distance of movement was longer than that of sitting and turning his/her face and body, the change of the appearance was clear, and it was suitable for measuring eye movements.

4.1.3. Procedures

Test. The participants wore the eye tracker and carried a backpack containing recording units. In all groups, 16 trials per block were conducted as with the BL phase of Experiments 1 and 2. The procedures such as the composition of the block were the same as those of BL.

Mutual gaze condition. This condition was conducted only in the ASD group. The experimenter presented the sample stimuli at 0°, 90°, 180°, or 270°. One of the eyes and the nose were glittering alternately for 0.75 s. Then, the experimenter instructed participants to “Walk while looking at the glittering eye and nose.” and the participants walked to another viewpoint. The nose was chosen because it was at the center of the facial stimulus and the spatial relation with the eyes was most obvious. A black slide was inserted so participants did not pay attention to the monitor when returning to self-viewpoint. Immediately after the participants returned, the experimenter presented the comparison stimuli, and the participants chose which eye was glittering (**Figure 2**). While moving to another viewpoint, the experimenter monitored the eye movements and provided feedback such as “You looked at the eye and nose well.”

4.1.4. Dependent Variables and Data Analysis

The dependent variables were 1) *number of correct responses*, 2) *frequency of viewpoint movement behaviors*, and 3) *percentage of fixation duration in each part of the sample stimuli* (%). The number of correct responses and frequency of viewpoint movement behaviors were analyzed similarly to Experiments 1 and 2. To measure fixation duration, eye movement analysis software (Tobii technology Tobii Pro Lab) was used and areas of interest (AOI) were manually defined: hair, glittering eye, non-glittering eye, nose, and mouth areas. The AOI

total fixation duration(s), not including the saccade, was automatically calculated for each sample stimuli. The total time each participant had fixed each AOI was divided by the total time of all AOIs and multiplied by 100. In addition, the mean values of 1) to 3) for each group were calculated from these data.

4.2. Results

The FU and TD groups had similar numbers of correct responses, and the FU group had a higher frequency of viewpoint movement behaviors than that of the TD group (Figure 5). In the test, the line of sight in the FU and TD groups tended to fixate on the hair, nose, and mouth areas, with the glittering eye area as the center, regardless of whether the viewpoint movement behavior occurred. However, the percentage of fixation duration in the glittering eye area was slightly higher in the FU group than in the TD group at all angles. In the test, the line of sight in the FU and TD groups tended to fixate on the entire facial figure, whereas that in the ASD group tended to fixate mainly on the glittering eye area. In the mutual gaze condition, the line of sight focused mainly on the glittering eye and nose areas. The results of each group are described below.

First, in the FU group, Fuku's number of correct responses was 3 times at 180° and correct responses occurred in all trials at other angles (Figure 6). For other participants, it was 4 times at all angles. In all participants, viewpoint movement behaviors occurred in all trials at 90°, 180°, and 270°. The percentage of fixation

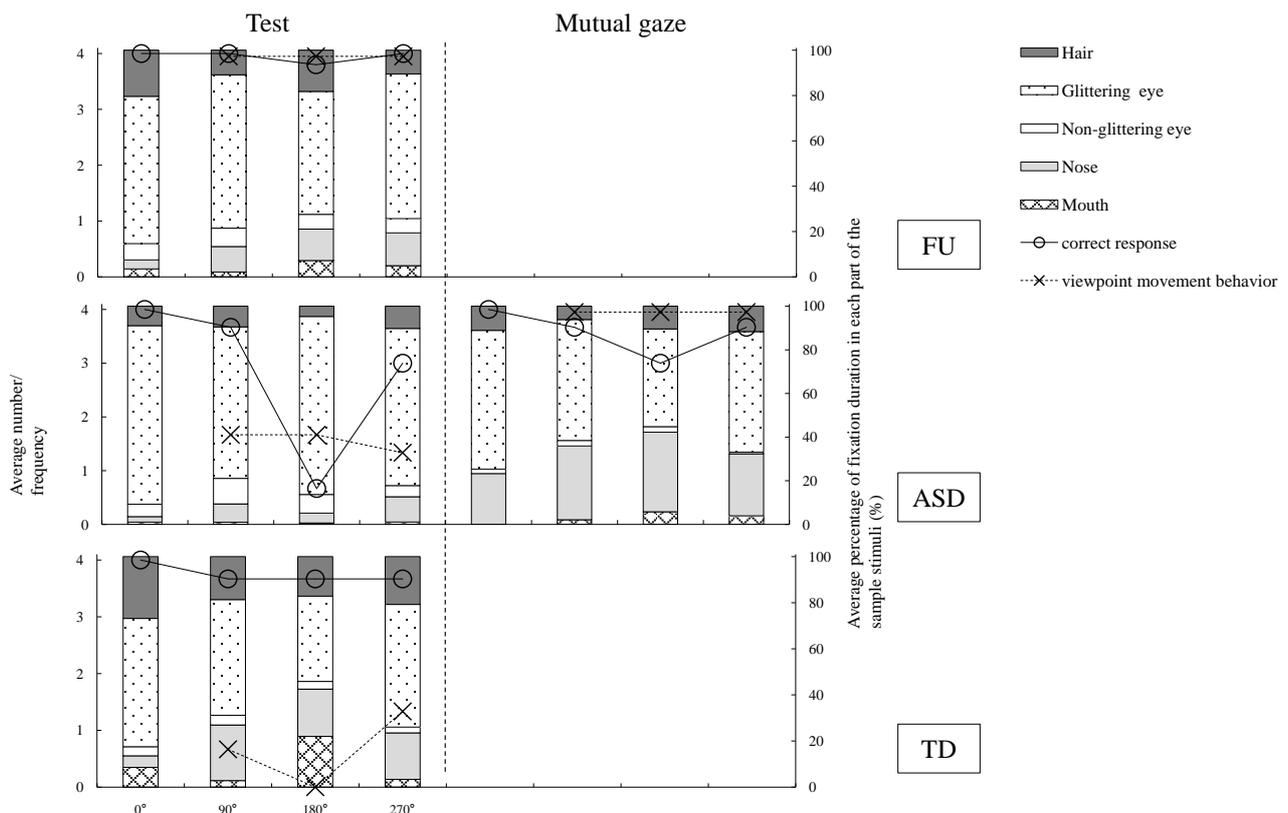


Figure 5. Results of each group in Experiment 3.

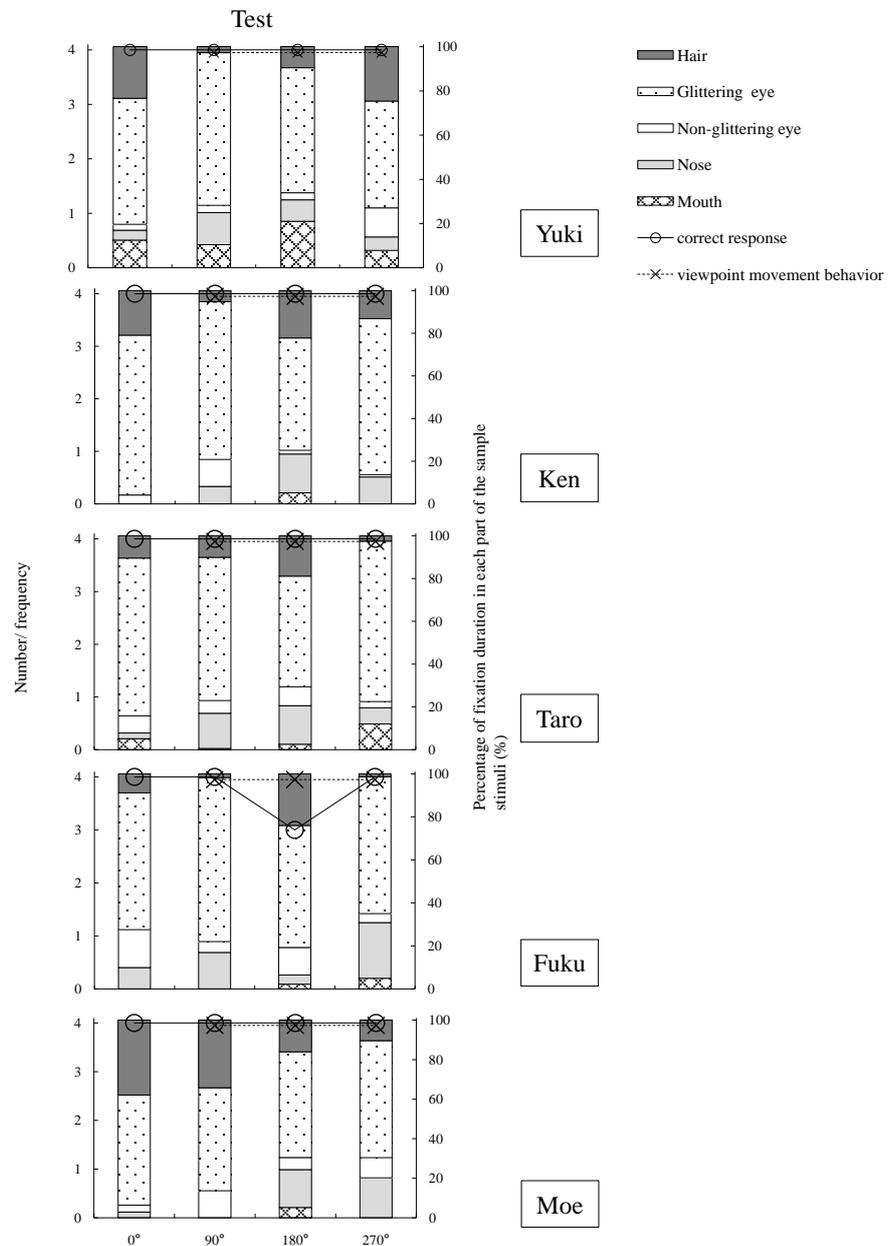


Figure 6. Results of the FU group in Experiment 3.

duration in the glittering eye area at 180° was lower than the other angles and the percentage in the other areas increased for all participants. Second, regarding the number of correct responses in the test of the ASD group, correct responses occurred in all trials at 0° for all three participants, and those of 90°, 180°, and 270° were 3 to 4, 0 to 1, and 2 to 4 times, respectively (Figure 7). The viewpoint movement behavior was 0 times for Sora, which occurred in most trials with Ryo, and occurred once for each angle with Hina. The percentage of fixation duration in the glittering eye area was high for all three participants. In the subsequent mutual gaze condition, the number of correct responses of all participants tended to increase, and the percentage of fixation duration in the nose area

mainly increased. Additionally, the fixation duration in the mouth area adjacent to the nose area slightly increased compared to the test. Third, in the TD group, Jun's number of correct responses at 0° was 4 times and 3 times at the other angles (Figure 8). Gaku and Anna's correct responses were 4 times at all angles. The viewpoint movement behavior did not occur with Jun and Anna. It occurred 2 to 4 times at each angle with Gaku. Similar to the FU group, the percentage of fixation duration in the glittering eye area decreased at 180°, and the percentage in the nose area in Jun, the mouth area in Gaku, and the hair area in Anna increased.

4.3. Discussion

In Experiment 3, we analyzed the eye movements when sample stimuli were presented and verified the effects of the condition that guided the line of sight in the spatial array between the components. As a result, the line of sight in the participants of the FU and TD groups tended to fixate on the entire facial figure in the ASD group with the eye area especially at 180°. In the mutual gaze condition, the line of sight in participants in the ASD group mainly fixated on the eyes and nose areas, and the number of correct responses increased. Therefore, the analysis of eye movements demonstrated the effectiveness of the conditions suggested by Experiments 1 and 2.

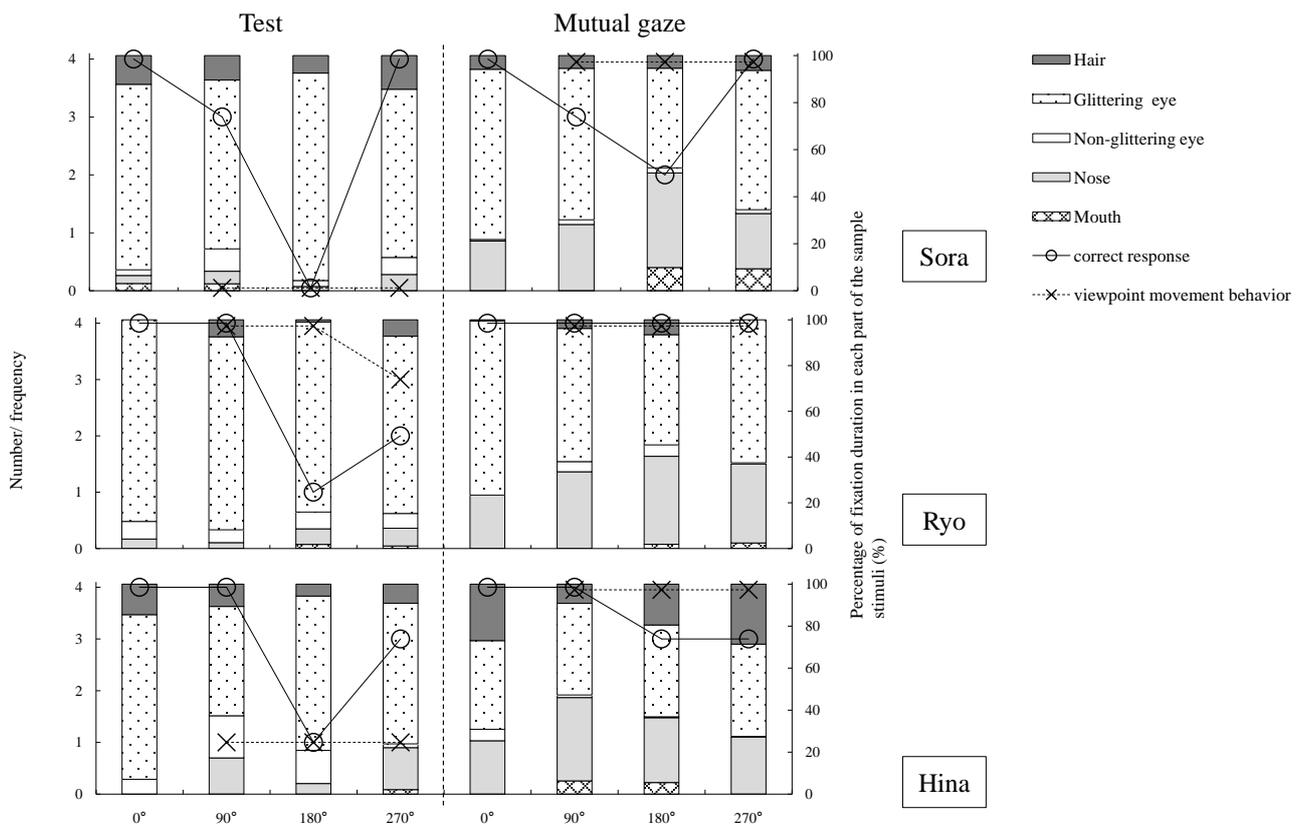


Figure 7. Results of the ASD group in Experiment 3.

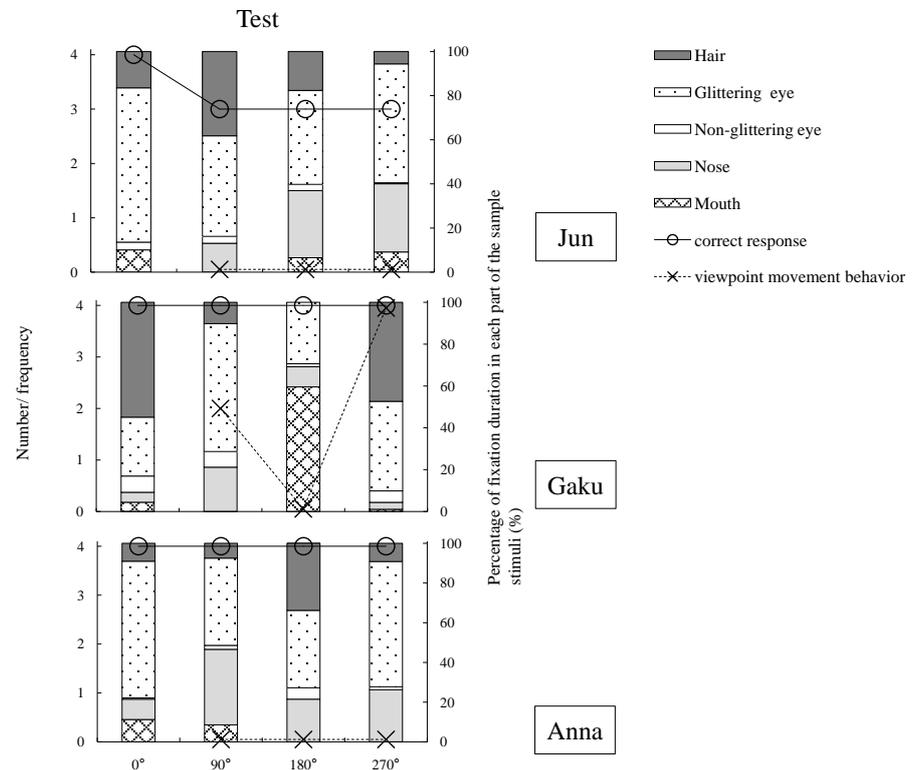


Figure 8. Results of the TD group in Experiment 3.

The eye trackers have not been used in previous studies of VPT (Pearson et al., 2013). In conventional eye tracking, an eye tracker installed near a monitor is mainly used to analyze where a participant is looking on the monitor. This method requires that the distance between the participant's head and the eye tracker be kept somewhat constant. In VPT2, the sensation and minute movement of the body is in an unstable position (Watanabe, 2016), and physically turning or moving the body to another viewpoint (Asaoka et al., 2019; Kessler & Thomson, 2010) affects performance. Thus, VPT2 is very sensitive to body sensation and position. In Experiment 3, the fixation duration was displayed during the performance of the VPT2 task in children with ASD by using the most recent technology called real-world eye tracking. In addition, the results that participants of the FD group moved while mutually gazing at the eyes and other areas suggests that the training conditions might not only memorize and respond to the patterns (Falla & Alós, 2016; Okuyama & Isawa, 2010) but also promote the coding of spatial arrays. This is also supported by the results of test and mutual gaze conditions in the participants of the ASD group. During the test, the fixation duration in the eye area was relatively long. These results mean that the participants in the ASD group looked at the facial figure but were less likely to focus on the positional relations between the eyes, that is, the frequency of mutual gaze behaviors is low. In the mutual gaze condition, the gaze behaviors toward the eyes and nose areas was prompted, the coding was promoted, and the number of correct responses increased. Furthermore, in Jun and Anna of the TD group, the

viewpoint movements did not occur, but generally correct responses occurred, and the result of looking at the entire facial figure might show evidence of mentally shifted perspectives.

Experiment 3 has some limitations. First, the eye tracker should be used in the same method as Experiment 2, to follow the changes of eye movements and gaze time during movement. Second, it is necessary to guarantee the representativeness of results by increasing the number of participants and perform a statistical analysis to show significant differences within and between the groups. Third, the VPT1/VPT2 assessment (e.g., Hadwin, Howlin, & Baron-Cohen, 2015) should be conducted to assess the development of the VPT before and after the introduction of the conditions.

5. General Discussion

In the present study, we implemented the Face Rotation Task (Watanabe, 2000) and hypothesized that retaining the visual representation from another viewpoint and coding the relation between components in a spatial array (Huttenlocher & Presson, 1979) would facilitate the establishment of VPT2. In Experiment 1, we examined the factors of TL that occur from the removal of the sample stimuli to the presentation of the comparison stimuli and suggested that setting the TL according to the cases is a prerequisite for the establishment of VPT2. Additionally, the importance of moving to another viewpoint while continuing to look at sample stimuli was indicated. These results suggest that memory retention and perspective transformations are interrelated. Thus, in Experiment 2, we examined the effect of setting of appropriate TL and shaping gaze behaviors on the sample stimuli. As a result, the participants achieved the criteria of accomplishment with a smaller number of blocks than Experiment 1. Furthermore, in Experiment 3, we analyzed eye movements during the presentation of the sample stimuli and confirmed the effectiveness of the conditions suggested by two experiments. From the above, we can deduce that 1) the participants move from the self to another viewpoint while continuing to look at the stimuli and 2) they move from another to self-viewpoint while retaining the representation from another viewpoint.

Previous studies have demonstrated that inactivation of perspective transformations has an impact on the difficulty of guessing the appearance from another viewpoint in children with ASD (Conson et al., 2016; Conson et al., 2013). This study partially expanded the findings of previous studies in that it clarified some conditions that promote the establishment of VPT2. We consider the possibility that the observation of the appearance from various viewpoints and memory retention are interconnected in children with ASD who have not sufficiently acquired VPT2 skills.

The average gaze time during movement as the number of correct responses increased in Experiments 1 and 2. Why does moving to another viewpoint while continuing to look at sample stimuli encourage its acquisition? In other words, why was the number of correct responses low when the participants looked at

the facial stimuli in the upright position *only* from another viewpoint? In answering this question, we rearranged processes of viewpoint movement. When the participants move to another viewpoint while looking at the stimuli, they observe the facial stimuli as a video, which becomes upright as they approach another viewpoint. On the other hand, if they look at the stimuli only from another viewpoint, they observe it in the upright position like a photograph. Especially in the case of 180°, the absolute spatial position of each part included in the sample and comparison stimuli is the same as that of the eyes, while hair, nose and mouth are inverted. For that reason, when presenting the comparison stimuli, the effort of decoupling “what I see now” (Hamilton et al., 2009) becomes higher than 90° and 270°. Moreover, the effort of the mental simulation of a body movement increases because the mental distance increases with increasing angles up to 180° clockwise or counterclockwise (Kessler & Rutherford, 2010). If it is difficult, the visual representation remains rotated, that is, only its *absolute* spatial position is retained (Watanabe, 2000). The *relative* spatial positions of the eyes and other parts are not held constant, making it difficult to provide correct responses. This result supports Pearson et al. (2014), who demonstrated that the difficulty of using the self as a reference frame has an impact on the difficulty of inferring the appearance of another viewpoint. In Experiment 3, the participants in the FU group moved to another viewpoint while looking at each other in the eye and other areas. These results suggest that coding spatial arrays included in stimuli from various viewpoints may contribute to decoupling in children with ASD who have not acquired VPT2 skills. When observing the sample stimuli from a location except from another viewpoint, the participants always perceived facial stimuli in a *misaligned* position from the upright position. In this process, they may learn that the positional relationship between the eye and other parts of the face is invariant from any viewpoint. Then, when the comparison stimuli are presented after returning to the original position, they discriminate that the current appearance is out of alignment with that of the observation. Thereby, they consider that the spatial relation of the sample stimuli is recalled, and the correct response occurs. From these task analyses, it is clear that discrimination precedes the recall of memory, however the significance of this study is to demonstrate that setting appropriate TL promotes the establishment of VPT2. Therefore, results indicate that training conditions in this study promote the response based on the pattern of the spatial array of elements included in a stimulus.

Future studies should verify task generalization of other VPT2 tasks, in addition to examining the issues listed in the discussion section of each experiment. Because even young children can understand the direction of the face (Carey, 1996), it is possible to understand the direction and spatial relation of the included elements without performing VPT2 skills (Watanabe, 2000). Therefore, it is essential to confirm task generalization. It is also necessary to examine the relationship between cognitive and affective perspective-taking. Since VPT plays a fundamental role in promoting the development of perspective-taking (Shelton, Clements-Stephens, Lam, Pak, & Murray, 2012), future studies should demon-

strate the impact of the acquisition of VPT2 skills on social development. This may also lead to increased educational significance of early intervention for children with ASD. Thus, perspective-taking is positioned as an aspect of sociality and the research results may provide clues to support children with ASD. Future studies should consider its application to the practical implications of understanding relationships with others, such as social skills, imitation, and understanding of words such as “go/come” and “give/receive.”

Acknowledgements

The authors are very grateful to all the children and their parents who participated in this study. We would also like to thank M. Nagai for her assistance with the experiments and Y. Choi for lending the eye tracker.

Funding

This study was supported by a Grant-in-Aid for Japan Society for Promotion of Science (JSPS) Fellows (No. 16J00464).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- American Psychiatric Association (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). Washington, DC: American Psychiatric Association.
<https://doi.org/10.1176/appi.books.9780890425596>
- Asaoka, H., Kumagai, M., Okamura, S., & Watanabe, M. (2016). Effects of Observing the View from Another Person’s Viewpoint on Training of Spontaneous Spatial Perspective-Taking in a Child with Autism Spectrum Disorder. *The Japanese Journal of Autistic Spectrum*, 14, 5-11.
- Asaoka, H., Takahashi, T., Chen, J., Fujiwara, A., Watanabe, M., & Noro, F. (2019). Difficulties in Spontaneously Performing Level 2 Perspective-Taking Skills in Children with Autism Spectrum Disorder. *Advances in Autism*, 5, 243-254.
<https://doi.org/10.1108/AIA-09-2018-0028>
- Avikainen, S., Wohlschläger, A., Liuhanen, S., Hänninen, R., & Hari, R. (2003). Impaired Mirror-Image Imitation in Asperger and High-Functioning Autistic Subjects. *Current biology*, 13, 339-341. [https://doi.org/10.1016/S0960-9822\(03\)00087-3](https://doi.org/10.1016/S0960-9822(03)00087-3)
- Ayres, A. J. (1979). *Sensory Integration and the Child*. Los Angeles, CA: Western Psychological Services.
- Baron-Cohen, S. (1989). Perceptual Role Taking and Protodeclarative Pointing in Autism. *British Journal of Developmental Psychology*, 7, 113-127.
<https://doi.org/10.1111/j.2044-835X.1989.tb00793.x>
- Carey, S. (1996). Perceptual Classification and Expertise. In R. Gelman, & T. Au (Eds.), *Perceptual and Cognitive Development* (pp. 49-69). San Diego, CA: Academic Press.
<https://doi.org/10.1016/B978-012279660-9/50020-8>
- Conson, M., Hamilton, A., De Bellis, F., Errico, D., Improta, I., Mazza-Rella, E., et al.

- (2016). Body Constraints on Motor Simulation in Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*, *46*, 1051-1060. <https://doi.org/10.1007/s10803-015-2652-x>
- Conson, M., Mazzarella, E., Froli, A., Esposito, D., Marino, N., Trojano, L., et al. (2013). Motor Imagery in Asperger Syndrome: Testing Action Simulation by the Hand Laterality Task. *PLoS ONE*, *8*, e70734. <https://doi.org/10.1371/journal.pone.0070734>
- Elekes, F., Varga, M., & Király, I. (2017). Level-2 Perspectives Computed Quickly and Spontaneously: Evidence from Eight- to 9.5-Year-Old Children. *British Journal of Developmental Psychology*, *35*, 609-622. <https://doi.org/10.1111/bjdp.12201>
- Falla, D., & Alós, F. J. (2016). Contextual Control in Visuospatial Perspective-Taking Skills in Adults with Intellectual Disabilities. *Behavioral Interventions*, *31*, 44-61. <https://doi.org/10.1002/bin.1435>
- Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young Children's Knowledge about Visual Perception: Further Evidence for the Level 1-Level 2 Distinction. *Developmental Psychology*, *17*, 99-103. <https://doi.org/10.1037/0012-1649.17.1.99>
- Hadwin, J. A., Howlin, P., & Baron-Cohen, S. (2015). *Teaching Children with Autism to Mind-Read: The Workbook*. Chichester: John Wiley & Sons.
- Hamilton, A. F., Brindley, R., & Frith, U. (2009). Visual Perspective Taking Impairment in Children with Autistic Spectrum Disorder. *Cognition*, *113*, 37-44. <https://doi.org/10.1016/j.cognition.2009.07.007>
- Hobson, R. P. (1984). Early Childhood Autism and the Question of Egocentrism. *Journal of Autism and Developmental Disorders*, *14*, 85-104. <https://doi.org/10.1007/BF02408558>
- Hobson, R. P., & Hobson, J. A. (2008). Dissociable Aspects of Imitation: A Study in Autism. *Journal of Experimental Child Psychology*, *101*, 170-185. <https://doi.org/10.1016/j.jecp.2008.04.007>
- Huttenlocher, J., & Presson, C. C. (1979). The Coding and Transformation of Spatial Information. *Cognitive Psychology*, *11*, 375-394. [https://doi.org/10.1016/0010-0285\(79\)90017-3](https://doi.org/10.1016/0010-0285(79)90017-3)
- Ito, H., Tani, I., Yukihiro, R., Adachi, J., Hara, K., Ogasawara, M., et al. (2012). Validation of an Interview-Based Rating Scale Developed in Japan for Pervasive Developmental Disorders. *Research in Autism Spectrum Disorders*, *6*, 1265-1272. <https://doi.org/10.1016/j.rasd.2012.04.002>
- Kessler, K., & Rutherford, H. (2010). The Two Forms of Visuo-Spatial Perspective Taking Are Differently Embodied and Subserve Different Spatial Prepositions. *Frontiers in Psychology*, *1*, 213. <https://doi.org/10.3389/fpsyg.2010.00213>
- Kessler, K., & Thomson, L. A. (2010). The Embodied Nature of Spatial Perspective Taking: Embodied Transformation versus Sensorimotor Interference. *Cognition*, *114*, 72-88. <https://doi.org/10.1016/j.cognition.2009.08.015>
- Ledford, J. R., & Gast, D. L. (2018). *Single Case Research Methodology: Applications in Special Education and Behavioral Science* (3rd ed.). London: Routledge. <https://doi.org/10.4324/9781315150666>
- Leekam, S., Baron-Cohen, S., Perrett, D., Milders, M., & Brown, S. (1997). Eye-Direction Detection: A Dissociation between Geometric and Joint Attention Skills in Autism. *British Journal of Developmental Psychology*, *15*, 77-95. <https://doi.org/10.1111/j.2044-835X.1997.tb00726.x>
- Leslie, A. M., & Frith, U. (1988). Autistic Children's Understanding of Seeing, Knowing and Believing. *British Journal of Developmental Psychology*, *6*, 315-324.

<https://doi.org/10.1111/j.2044-835X.1988.tb01104.x>

- Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A Revised Version of a Diagnostic Interview for Caregivers of Individuals with Possible Pervasive Developmental Disorders. *Journal of Autism and Developmental Disorders*, 24, 659-685. <https://doi.org/10.1007/BF02172145>
- Meyer, J. A., & Hobson, R. P. (2004). Orientation in Relation to Self and Other: The Case of Autism. *Interaction Studies*, 5, 221-244. <https://doi.org/10.1075/is.5.2.04mey>
- Montefinese, M., Sulpizio, V., Galati, G., & Committeri, G. (2015). Age-Related Effects on Spatial Memory across Viewpoint Changes Relative to Different Reference Frames. *Psychological Research*, 79, 687-697. <https://doi.org/10.1007/s00426-014-0598-9>
- Morss, J. R. (1987). The Construction of Perspectives: Piaget's Alternative to Spatial Ego-centrism. *International Journal of Behavioral Development*, 10, 263-279. <https://doi.org/10.1177/016502548701000301>
- Ohta, M. (1987). Cognitive Disorders of Infantile Autism: A Study Employing the WISC, Spatial Relationship Conceptualization, and Gesture Imitations. *Journal of Autism and Developmental Disorders*, 17, 45-62. <https://doi.org/10.1007/BF01487259>
- Okuyama, T., & Isawa, S. (2010). Right-Left Discrimination from One's Own and Another Person's Viewpoint in Children with Autism: Higher-Order Conditional Discrimination and Generalization of Viewpoint. *Japanese Journal of Behavior Analysis*, 24, 2-16.
- Parent-Interview ASD Rating Scale (PARS) Committee (2013). *Oya mensetsu shiki Jiheisupekutoramusyo Hyotei Syakudo Tekisuto Kaiteiban [Parent-Interview ASD Rating Scale-Text Revision]*. Tokyo: Spectrum Publishing Co.
- Pearson, A., Marsh, L., Hamilton, A., & Ropar, D. (2014). Spatial Transformations of Bodies and Objects in Adults with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, 44, 2277-2289. <https://doi.org/10.1007/s10803-014-2098-6>
- Pearson, A., Ropar, D., & Hamilton, A. F. D. (2013). A Review of Visual Perspective Taking in Autism Spectrum Disorder. *Frontiers in Human Neuroscience*, 7, 652. <https://doi.org/10.3389/fnhum.2013.00652>
- Reed, T., & Peterson, C. (1990). A Comparative Study of Autistic Subjects' Performance at Two Levels of Visual and Cognitive Perspective Taking. *Journal of Autism and Developmental Disorders*, 20, 555-567. <https://doi.org/10.1007/BF02216060>
- Shelton, A. L., Clements-Stephens, A. M., Lam, W. Y., Pak, D. M., & Murray, A. J. (2012). Should Social Savvy Equal Good Spatial Skills? The Interaction of Social Skills with Spatial Perspective Taking. *Journal of Experimental Psychology: General*, 141, 199-205. <https://doi.org/10.1037/a0024617>
- Slone, L. K., Smith, L. B., & Yu, C. (2019). Self-Generated Variability in Object Images Predicts Vocabulary Growth. *Developmental Science*, 22, e12816. <https://doi.org/10.1111/desc.12816>
- Smith, I. M. (1998). Gesture Imitation in Autism I: Nonsymbolic Postures and Sequences. *Cognitive Neuropsychology*, 15, 747-770. <https://doi.org/10.1080/026432998381087>
- Takano, Y. (1998). Why Does a Mirror Image Look Left-Right Reversed? A Hypothesis of Multiple Processes. *Psychonomic Bulletin & Review*, 5, 37-55. <https://doi.org/10.3758/BF03209456>
- Tanaka Institute for Educational Research (2003). *Tanaka Binet Chinou Kensa V [Tanaka-Binet Intelligence Scale V]*. Tokyo: Taken.
- Ueno, K., Nadeo, T., & Iinaga, K. (2008). *PVT-R Kaiga Goi Hattatsu Kensa [Picture Vo-*

cabulary Test-Revised]. Tokyo: Nihon Bunka Kagakusha.

Watanabe, M. (2000). Can Three-Year Old Children Take Spatial Perspectives? An Approach with the Face Rotation Task. *The Japanese Journal of Psychology*, *71*, 26-33.

Watanabe, M. (2016). Developmental Changes in the Embodied Self of Spatial Perspective Taking. *British Journal of Developmental Psychology*, *34*, 212-225.
<https://doi.org/10.1111/bjdp.12126>

Yirmiya, N., Sigman, M., & Zacks, D. (1994). Perceptual Perspective-Taking and Seriation Abilities in High-Functioning Children with Autism. *Development and Psychopathology*, *6*, 263-272. <https://doi.org/10.1017/S0954579400004570>

Yu, A., & Zacks, J. (2010). The Role of Animacy in Spatial Transformations. *Memory & Cognition*, *38*, 982-993. <https://doi.org/10.3758/MC.38.7.982>