4-4-2022

Manipulating Image Luminance to Improve Eye Gaze and Verbal Behavior in Autistic Children

LouAnne Boyd
Chapman University, lboyd@chapman.edu

Vincent Berardi
Chapman University, berardi@chapman.edu

Deanna Hughes
Chapman University, dhughes@chapman.edu

Franceli L. Cibrian
Chapman University, cibrian@chapman.edu

Jazette Johnson
University of California, Irvine

Recommended Citation

Manipulating Image Luminance to Improve Eye Gaze and Verbal Behavior in Autistic Children

Authors

LouAnne Boyd, Vincent Berardi, Deanna Hughes, Franceli L. Cibrian, Jazette Johnson, Viseth Sean, Eliza DelPizzo-Cheng, Brandon Mackin, Ayra Tusneem, Riya Mody, Sara Jones, and Karen Lotich

ARTICLE

Check for updates

Manipulating image luminance to improve eye gaze and verbal behavior in autistic children

LouAnne Boyd[1✉], Vincent Berardi[1], Deanna Hughes[1], Franceli Cibrian[1], Jazette Johnson[2], Viseth Sean[1], Eliza DelPizzo-Cheng[3], Brandon Mackin[1], Ayra Tusneem[1], Riya Mody[1], Sara Jones[4] & Karen Lotich[4]

Autism has been characterized by a tendency to attend to the local visual details over surveying an image to understand the gist–a phenomenon called local interference. This sensory processing trait has been found to negatively impact social communication. Although much work has been conducted to understand these traits, little to no work has been conducted to intervene to provide support for local interference. Additionally, recent understanding of autism now introduces the core role of sensory processing and its impact on social communication. However, no interventions to the end of our knowledge have been explored to leverage this relationship. This work builds on the connection between visual attention and semantic representation in autistic children. In this work, we ask the following research questions: RQ1: *Does manipulating image characteristics of luminance and spatial frequency increase likelihood of fixations in hot spots (Areas of Interest) for autistic children?* RQ2: *Does manipulating low-level image characteristics of luminance and spatial frequency increase the likelihood of global verbal responses for autistic children?* We sought to manipulate visual attention as measured by eye gaze fixations and semantic representation of verbal response to the question "What is this picture about?". We explore digital strategies to offload low-level, sensory processing of global features via digital filtering. In this work, we designed a global filter to reduce image characteristics found to be distracting for autistic people and compared baseline images to featured images in 11 autistic children. Participants saw counterbalanced images way over 2 sessions. Eye gaze in areas of interest and verbal responses were collected and analyzed. We found that luminance in non-salient areas impacted both eye gaze and verbal responding–however in opposite ways (however versus high levels of luminance). Additionally, the interaction of luminance and spatial frequency in areas of interest is also significant. This is the first empirical study in designing an assistive technology aimed to augment global processing that occurs at a sensory-processing and social-communication level. Contributions of this work include empirical findings regarding the quantification of local interference in images of natural scenes for autistic children in real-world settings; digital methods to offload global visual processing to make this information more accessible via insight on the role of luminance and spatial frequency in visual perception of and semantic representation in images of natural scenes.

[1] Chapman University, Orange, USA. [2] University of California, Irvine, USA. [3] Endicott College, Beverly, USA. [4] Speech and Language Development Center, Buena Park, USA. ✉email: lboyd@chapman.edu

Decades of research across several disciplines have considered differences in global processing as a core feature of autism. Across various processing systems, such as visual attention, sensory perception, and cognitive processing style, scholars have reported hierarchical differences between autistic and typical development. Atypical global processing has been attributed to local interference (Rinehart et al. 2000; Song and Hakoda 2015), weak central coherence (Happé et al. 2006; Jolliffe and Baron-Cohen 1999), enhanced perceptual functioning and the spatial frequency hypothesis (de Jong et al. 2008) or differential use of spatial frequency (Koh et al. 2010). Essentially, each theory describes local interference and attributes varying mechanisms that range from stimulus saliency (e.g., stimulus control) (Baisa et al. 2018, 2020; Guy et al. 2016) to executive functioning (e.g., attention control) (Fan and Posner 2004). As findings within and across mechanisms remain mixed across studies, researchers tend to remain focused on basic science rather than directly impacting the lives of autistic people (Doehring 2021). However, there is an opportunity to support the global processing via visual attention. Visual attention is linked to difficulties with social processing. Also, poor global processing can lead to semantic communication (Jolliffe and Baron-Cohen, 1999) as well as social challenges (Chawarska and Shic 2009; Deruelle et al. 2004; Kihara and Takeda 2019; Robertson and Baron-Cohen 2017), we hypothesize that filtering a scene to augment salient features could serve as an assistive technology—thus empowering autistic children access to global visual information in the real world.

Therefore, in this work, we combine the current understanding of local interference in autism research on several levels (e.g., vision, perception, cognition, neurology, speech, and language). By synthesizing findings related to local interference, we explore the feasibility of using a digital filter to shift one's gaze path to global features. To measure changes in visual attention, we collected verbal responses and eye movements.

Eye movements have been considered "a potential porthole into… cognitive processes" (Toet 2006). Yet, eye movement via eye trackers needs to be *confirmable* as the gaze behavior of attending is not readily observable (Dube et al. 2010), nor is *attending* by looking at a stimuli a guarantee that the stimuli has been processed (Liechty et al. 2003; Toet 2006). To address this gap, we designed the *think-see-say paradigm* that included a verbal response to capture the semantic representation of an image as well as eye tracking as measures of global processing. All three phases culminate in a real-word use case to support global processing by working-around local interference.

Think phase

Thinking requires executive functioning which is challenged in autism and is a core symptom of autism (Fan and Posner 2004). Executive functioning is both automatic and purposeful. Automaticity and effort are combined in such a way that allows for adaptive responses to novel complex situations across the day Happé et al. 2006). Within these executive functions, there is *cognitive control* (e.g., set-shifting, inhibition of responses) and *stimulus control* (e.g., materials, time) (Fan and Posner 2004; Happé et al. 2006). We leverage both concepts in this work as we prime participants to assist with *cognitive control*.

Priming is an important component of cognitive control but does not direct visual attention completely. Prompting for a specific level of processing improves perception of that level (Baisa et al. 2018), although cannot completely account for participants' attentional control to suppress other features (Baisa et al. 2018). For example, it is common in autism and developmental disorders to display overactive attention or overselectivity,

(Dickson et al. 2006; Ploog 2010; Rieth et al. 2015). Some behavioral interventions involve carefully prompting and shaping the environment to guide overactive attention (e.g., Pivotal Response Training, (Koegel and Wilhelm 1973)). To support cognitive control, we prompted participants to verbally answer the question "what is this picture about?" to prime for global responses.

See phase

To address stimulus control, we manipulate the image characteristics. We base our global filter intervention on research that described *local interference* as a key phenomenon in autism (Gargaro et al. 2018; Katagiri et al. 2013; Rinehart et al. 2000; Robertson and Baron-Cohen 2017; Song and Hakoda 2015) that could be "worked around" as local processing is intact. Local processing may become backlogged in the higher order processing when increased complexity is required to form a gist (Bertone et al. 2005; Grinter et al. 2010; Guy et al. 2016; Robertson and Baron-Cohen 2017). Other visual processes are also impacted in autism as well such as temporal binding (Robertson and Baron-Cohen 2017) and processing biological motion, which has been deemed a potential biomarker for autism (Kaliukhovich et al. 2020). However, temporal difficulties are a level of complexity beyond this fundamental work. We target intervention at static stimuli as there is still much debate over local interference in static experimental stimuli (Baisa et al. 2020; Chawarska and Shic 2009; Courchesne and Pierce 2005; Gargaro et al. 2018; Grinter et al. 2009; Gross 2005; Guy et al. 2016; Jolliffe and Baron-Cohen 1999; Katagiri et al. 2013; Koldewyn et al. 2013; Ploog 2010; Rieth et al. 2015; Rinehart et al. 2000; Song and Hakoda 2015). However, we examine visual attention to images of scenes of the natural world as a step beyond experimental stimuli, where little work has been conducted in autism. Natural images provide ecological validity, a critical consideration for application of insights into the real world (Wang et al. 2015). Specifically, we examine the relationship between image characteristics that make up the semantic features of an image using semantic content, luminance, and spatial frequency.

Spatial frequency relates to social cognition as it, in part, drives global processing (de Jong et al. 2008). Higher spatial frequency has been associated with local detail processing whereas lower spatial frequency support form and pattern perception (Ellemberg et al. 1999; Pasternak and Merigan, 1981), however both high and low spatial frequencies code structure of an image (Shulman et al. 1986; Shulman and Wilson 1987). Yet most natural images are dominated by low spatial frequency (Hughes et al. 1996). Subsequently, it is important to support global processing as autistic people may use spatial frequency differently (Koh et al. 2010).

Given this nuanced finding that low level processing is both intact and used preferentially, it is feasible that global processing could be prioritized (Baisa et al. 2018). We hypothesize that manipulating spatial frequency and luminance provide a preprocessing of global elements in the natural scene. Luminance like spatial frequency is perceived in dual streams (Badcock et al. 2005). Higher luminance draws the eye early in visual processing. Vision research confirms this and adds a third pathway that detects decrements to luminance (Badcock et al. 2005). Lastly, we consider the interaction of luminance and spatial frequency as spatial contrast sensitivity is dependent on luminance (Ellemberg et al. 1999).

Say phase: semantic representation

Verbal speech is a response that is based on the culmination of processing, language, and motor ability. Social communication is

Table 1 Participant demographics: age, social pragmatic score and details.

P#	Age	CELF-4	Specific speech related challenges
P1	18	1	Social referencing difficulties, perspective-taking difficulties
P2	18	2	Speech articulation difficulties, phonological challenges, grammar, and syntax difficulties
P3	19	1.5	Syntax/grammar difficulties (e.g., not sure when to use a pronoun), no ASD
P4	12	2	Delayed receptive and expressive language, perspective-taking difficulties
P5	16	2	Cognitive rigidity, perspective-taking difficulties, expressive language formulation difficulties, vocab difficulties: understanding and use
P6	12	2.5	Expressive formulation difficulties, word finding difficulty, perspective-taking difficulties
P7	12	1	Delayed response for language, expressive narrative language difficulty, language processing difficulties, perspective-taking difficulties – genuine in what he says but cannot provide rationale
P8	11	1	Cognitive rigidity, word finding difficulty, narrative language difficulty
P9	15	1.5	Delayed response time due in part to dysfluency, expressive semantic challenges
P10	9	2.5	Cognitive rigidity, (difficulty with transition – would not put on eye tracker), early sequencing difficulties, delayed response for language
P11	19	3	Delayed response time, language processing difficulties, pragmatic difficulties (difficulties with holding conversations, turn taking)

The CELF-4 rating is from 1–3 whereas 1 is mild and 3 is severe language disorder.

a pivotal skill (Koegel 2000) and is impacted by visual attention. For example, researchers have found that one's listening comprehension has been tied to tracking the same objects as a speaker (Richardson and Dale 2005). For example, if we know where the person is looking when speaking about a television program and we make these aspects visually brighter, we can improve the comprehension of the person who is listening to the speaker (Richardson and Dale 2005). When we perceive these natural scenes, global (big picture information) and local (details) are dependent on *lexical-semantic processing* (i.e., the forest helps us see the trees and the reverse is true) (Bouvet et al. 2011). Therefore, the manipulation of the visual stimuli could have an impact on the cognitive process (Richardson and Dale 2005), which may be observable via a verbal response. At the onset and after each stimulus presentation, we show a prompt that says "what was the picture about" to demonstrate the level of processing employed. The literature purports that autistic children demonstrate a preference for stimuli at the local level unless they are instructed to do otherwise (Koldewyn et al. 2013). Therefore, we instructed the participants to think globally by using the prompt, "What was the picture about?"

In summary, we aim to provide real-world strategies to promote access to global processing by working around the low-level characteristics related to local interference. We aim to integrate many of these fields and carefully consider the task demands in our study so that they will be useful for solving real world challenges for autistic people (Doehring 2021). To the best of our knowledge, this is the first empirical study in designing an assistive technology aimed to augment global processing using a sensory-first perspective of autism (Robertson and Baron-Cohen 2017). This work builds on the connection between visual attention and semantic representation in autistic children (Plesa Skwerer et al. 2019). Contributions of this work include empirical findings regarding applied research on quantifying local interference in autistic children in real-world settings and insight on the relationship between visual perception and semantic representation.

Our work aims to augment global processing by working around the local interference as visual preference and visual attention appear to remain locally-focused into adulthood (Kaliukhovich et al. 2020), so this work aims to determine if visual attention is changeable in individuals with ASD (Kaliukhovich et al. 2020). To answer this question, we manipulate visual attention as measured by eye gaze fixations and semantic representation of verbal response.

RQ1: Does manipulating image characteristics of luminance and spatial frequency increase likelihood of fixations in hot spots (Areas of Interest) for autistic children?

RQ2: Does manipulating low-level image characteristics of luminance and spatial frequency increase the likelihood of global verbal responses for autistic children?

Methods

Study setting and participants. The current study was conducted at a nonpublic school that specializes in speech and language disorders. All children who attend this school automatically participate in intervention programs that support language and social interactions. We recruited 11 children who receive highly specialized speech and language services related to social pragmatics and/or an autism diagnosis, ages 9–18. Ten of the eleven have a special education eligibility of diagnosis of Autism. P3 does not. They have a range of speech and language issues described in Table 1: demographics. Parents returned a signed consent form prior to the study, and children were asked to provide assent at the onset of each session. Two sessions were conducted across consecutive days to allow time between viewing of the image in both baseline and filtered conditions in the same sitting. We did not screen for local interference, however, previous research reports local interference to be likely (Gargaro et al. 2018; Kaliukhovich et al. 2020; Katagiri et al. 2013; Koldewyn et al. 2013; Rinehart et al. 2000), and all participants had difficulty with semantic and pragmatic language that are described in Table 1. We also solicited a customized description of each participant's social-pragmatic language ability as no standard scale of language that reflects global processing of our set of naturalistic images exists. We make the widely accepted assumption that eye gaze, a measure of one's overt attention (observable through eye tracking technology) reflects the participants' covert attention (Liechty et al. 2003).

Data collection

Collecting verbal response. An independent SLP created a scoring category unique for this set of images and sample population that answered, "What was the picture about". Over the course of 2 months, 7 SLPs (2 senior SLPs, 5 masters students SLPs) worked to achieve reliability using the scoring rubric. The final response categories contained a global description plus additional relevant information. The scoring rubric ultimately contained
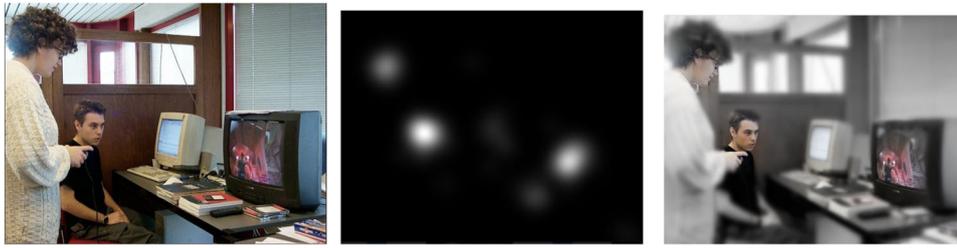
Fig. 1 Example of baseline and filtered image with heatmap of neurotypical viewing from OSIE open-source data set. Right image is a sample picture from the OSIE dataset in its raw form. The middle image is the heatmap of neurotypical viewing, the left image is the filtered version of the same image using the global filtered we created. This figure is not covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of LouAnne E. Boyd; copyright © LouAnne E Boyd, all rights reserved.

ordered categories with scores ranging from 0 to 2 with 0 = incorrect/unrelated responses; 1 = irrelevant or local details; and 2 = plausible global description. To be sure not to privilege verbal responses that were superior in terms of level of vocabulary or syntax use, the coders did not score responses regarding the assumed gender of people in the photos nor correctness of pre-positions, but simply the responses containing local details not relevant to the gist of the scene.

Collecting eye gaze behavior. Participants wore a head mounted eye tracker found at positivescience.com during the sessions. We chose a head mounted device because it could be utilized in real-world settings. Calibration to align aspects of the screen with the eye tracker occurred at the onset of every session (again, except for P8 who did not choose to wear the eye tracker in session 1). The calibration included gazing at five points on a calibration screen. Calibration took approximately 2 min per participant. Participants' point of view video from the head-mounted device was captured. The recording overlaid crosshairs on the image to indicate eye gaze location. The *think-see-say* paradigm was presented as an auto-advancing PowerPoint on a 27-inch monitor where the children sat approximately 23 inches from the screen. Each image was presented for 3 s followed by a 7 s screen that read "What was this picture about?" A familiar SLP sat next to the participant to maintain their engagement.

Study design. The experimental design of the study was a 2 × 2 within group factorial design (baseline × filter, session one × session two) where we randomized the presentation of 50 images in both their original form and counterbalanced with filtered versions. We presented across 2 days, so the same images were not viewed in the same sitting. When we explained the study, we walked through the task with three practice images and prompted them to tell us, "What the picture was about" after each image presentation. In the study, we played an automated PowerPoint that showed the image, text appeared on the screen, "What is the picture about?" We video recorded their verbal responses. Our hypothesis was that the eye gaze fixations would shift from locally-salient areas (high pixel contrast) to globally-salient areas (hot spots of the Areas of Interests in heat maps) in the filtered condition because of the manipulation in spatial frequency and luminance, and that this would lead to a shift in verbal response to the semantic priming prompt, "What is the picture about" to more global responses in the filtered condition as well.

Two sessions were conducted across consecutive days—except for P9 who completed one session in the morning and one in the afternoon. We employed simple randomization to determine for each of the 50 images which would appear first, (baseline or filtered). We purposely divided the study into two sessions to rule out any possible carryover from seeing the same image in the same sitting (in both baseline and filtered condition). Each

session was conducted by the participant's SLP. The SLP sat next to the child, provided the introductory explanation at the start of the session, and any redirection to look at the fixation drift point that was presented between stimuli. The slide presentation took 9 min per session. The cost of using the head mounted eye tracker was the lack of alignment with the screen which necessitated hand scoring. However, we believed the benefit resulted in improved ecological validity due to the natural setting.

Intervention: building the assistive technology
Given that machines are able to "see" objects in terms of objects or proto-objects (Yanulevskaya et al. 2013), and can also predict human eye gaze (Xu et al. 2014), we use these features to redirect the visual attention eye gaze patterns found in autism (Wang et al. 2015). Combining these concepts, we took inspiration to create an assistive technology that could automate "seeing the big picture" to prompt seeing the gist. Seeing the main components of an image leads to understanding the picture as a whole unit.

We filtered images by semantic saliency based on neurotypical viewing to ensure we captured areas of initial processing by humans rather than using a saliency filter algorithm (Perazzi et al. 2012). We lowered the spatial frequency by decreasing the pixel-level contrast.

In this work, we leverage previous research that provides the heatmaps of natural viewing eye gaze of neurotypical (NT) adults as a template for global features-as the first moment of viewing is considered to be a broad sweeping for global meaning (Spering and Carrasco 2015). To create the filter, we digitally-manipulate images by desaturating and blurring non-relevant detail, as determined by heat maps of eye gaze of neurotypicals (NTs) collected for a set of previously studied images, see Fig. 1. Desaturation occurred in the same fashion where for every point on the image; the corresponding point from the image's heatmap is used to determine the level of luminance to the relative degree.

Leveraging an open-source data set of natural images and heatmaps of neurotypical fixations. Natural viewing and viewing natural scenes (images of the physical world versus hierarchical figures) are important to consider as the low spatial frequency makes up most of natural terrain and textures and are statistically different from artificial stimuli such as random dot patterns (Burton and Moorhead 1987; Field 1987; Hughes et al. 1996). Human faces also fall into a low spatial frequency range (Hughes et al. 1996; Kihara and Takeda 2019) and often noted as processed differently in autism. Therefore, we used images of natural settings to support ecological validity.

The images come from an open-source repository with 700 images and their corresponding eye gaze heatmaps. The heatmaps were produced using a model based on the aggregate of 3 s of natural viewing by 20 NT adults (Xu et al. 2014). We used the first 50 images. We also used the heatmaps for these images that

are provided in the open-source repository. The previous researchers identified areas of interest at a semantic level in the baseline images. More specifically, we leverage the semantic areas predefined by (Wang et al. 2015), who describe three levels of visual attention based on eye fixations between groups as semantic (gist), object, and pixel. They found NTs focused more on semantic features while autistic participants focused more on pixel level features. We leverage these findings to develop a tool to guide visual attention.

Given we are interested in making assistive technology, we utilize real-world images. We not only examine natural images for their generalizability but also, we break down each stimulus into global and local parts, giving differential meaning beyond the totality of spatial frequency. Our aim is to direct *how to use* the salient features to guide eye gaze.

Data analysis

We conducted a general linear regression to understand the relationships between the two global processing measures of global verbal responses and eye fixations in global hot spots. The human factors model analyzed: age, degree of social pragmatic language impairment. The image characteristics model analyzed spatial frequency, luminance, semantic content, hot spot size, within-image differences in saliency. The study design model analyzed condition, session order, and item order.

Analyses of verbal responses. Participant responses to, "What was the picture about" were coded by two outside speech-language pathology students who scored the data independently and were unaware of the purpose of the study. All verbal responses were scored. The scores ranged from 0 to 2 with 0 = incorrect/unrelated responses; 1 = irrelevant or local details; and 2 = plausible global description. During the processing of the data, from the 550 possible response pairs, 75 pairs deleted because the participant said nothing for at least one of the trials in the baseline/filtered pair. The remaining 475 pairs were scored, interrater reliability which was found to be 87%.

Analysis of eye tracking. Eye tracking lends itself to this task as the initial eye gaze fixation of a scene is believed to be based on characteristics of both the viewer and task (i.e., top-down) as well as features of the stimulus (i.e., bottom-up). These local over global differences can be further understood by examining the function of fixations as less than 10% of the visual field is projected on the fovea (Wedel et al. 2008). Therefore, when the eye gaze is paused, we assume the fixation affords taking in the local detail (Campana et al. 2016; Vision 1985).

Additionally, rather than pre-define the socially salient areas (percent of fixation time of looking at the eyes versus the mouth of a face) as did (Klin et al. 2002), we targeted global salience as the AOI hot spots viewed by NT young adults in a naturalistic viewing session in previous work by (Wang et al. 2015) and scored hit or miss for 3 s of viewing. Following similar scoring procedures as (Klin et al. 2002) who also scored video from a head-mounted eye tracking with image and crosshairs, we enlisted six undergraduate research assistants to view the 848 three-second video clips at 0.25 playback speed. Our scorers viewed each video to determine if there was at least one hit or miss in the hot spots based on the eye gaze path passing through any hot spot in each image. The scoring yielded an interobserver agreement score of 82%.

Determining image characteristics. The characteristics of images that are discussed here include: luminance, size of the hot spots, and semantic content. Luminance is one attribute that is sensed

early in human visual processing. Therefore, this attribute is part of the early global impression of an image, making it the area to target for an intervention aimed to highlight global processing. Size of a hot spot may also have an impact on fixation in hot spots (e.g., hit rate) as a larger area that counts as a hit, provides more chances of a hit. Semantic content (i.e., what is going on in the picture in terms of meaning) is processed after the image has been sensed and is processed using attention–cognitive resources. We hypothesized that each of these characteristics could play an important role in forming a global filter because of their effects: the intensity of light, the size of the area of interest (i.e., hot spot), and the distribution of content inside the image. Lastly, the variation within the image of light and spatial frequency places a role, not just the average of it across the image.

Semantic content. The 50 pictures are categorized based on the content of the photo to capture the semantic elements. Initially three research assistants categorized the pictures into six categories which were: people, animals, objects, rooms, transportation, and food. The interobserver agreement results were not satisfactory as some images fell into multiple groups so the categories were further reduced to two categories described as living or nonliving. This coding scheme yielded reliable sorting and was added as a semantic-level variable in our analysis. Given that the images had numerous hot spots, we confirmed that each of the 32 images (64% of images) contained hot spots on all the living elements in the picture.

Size of the hot spot. As each image has unique placement of features and the size of the hot spots range, we calculated the total number of pixels that fell within the hot spots. To understand if the size of a hot spot zone had an impact on fixation in hot spots, we compared this number to the total pixels per image, 480,000 pixels, to determine the size of the hot spot per image. We used the luminance values from the OSIE data set of grayscale heatmaps (Wang et al. 2015). We divided the luminance scale (0–255) in half so that the pixels with luminance value at 125 or greater were labeled within a hot spot. Delineating the hot spot zones from the not-hot spot-zones allowed us to compare not only the impact of size but also the other image characteristics of luminance in this work.

Luminance. We explored luminance but did not add chroma to our analysis as luminance and chroma were highly correlated for our image set. The correlation between chroma and luminance was very high ($p = 0.99$) (Sean 2020). Nonetheless, we desaturated the images so that the chroma and brightness of non-global features are lowered because previous research has found the autistic participants tended to focus on bright contrast at the pixel level (Wang et al. 2015). To quantify the luminance of a given image, the pixels of each image were converted into values for RGB (Red, Green, Blue) and then converted into HLS (Hue, Lightness, Saturation) by utilizing a Python function. We used only the lightness value of each variable as for luminance.

Spatial frequency. For a given image, the overall activity level is measured by the spatial frequency of the image (Eskicioglu and Fisher 1995). Spatial frequency describes the periodic distributions of light and dark in an image. High spatial frequency refers to features such as sharp edges and fine details, whereas low spatial frequency refers to features such as global shape. Overall, the spatial frequency in filtered images is considerably lower than the baseline images. We totaled the hot spot areas as images had multiple hot spots.

Human factors: severity of social pragmatic language impairment, age, gender. We requested non-identifiable information regarding each participant's age, gender, and specifics of language impairment from the SLP team. As participants displayed a wide range in language ability and came from different school districts to this specialized setting, there was no standardized measure of language or IQ across participants; however, IQ has not been found to impact response times and accuracy on Navon tests that contain clear global and local features (Guy et al. 2016). The SLPs scored each participant on their level of semantic language impairment based on subtests of the CELF-4 (Clinical Evaluation of Language Fundamentals-Fourth Edition) Pragmatics Profile. The areas it addressed were Rituals and Conversational Skills, Asking For, Giving, and Responding to Information and Nonverbal Communication Skills. These labels ranged from mild to severe and appear in Table 1 as part of the description of participants. Pertinent details of each participant to provide richer detail, see Table 1.

Results

RQ1: Does manipulating image characteristics of luminance and spatial frequency increase likelihood of fixations in hot spots (Areas of Interest) for autistic children?

RQ2: Does manipulating low-level image characteristics of luminance and spatial frequency increase the likelihood of global verbal responses for autistic children?

We answered our research questions by using a general linear model regression. Regarding RQ1, we found not only did specific image characteristics increase the likelihood of fixations in hot spots, so did participants and study design characteristics. Additionally, in relation to RQ2, we found that low-level image characteristics of luminance and spatial frequency increased the likelihood of global verbal responses, as did participant characteristics. See Table 2 for statistical results by all variables.

First, we considered the significant impact of the individual participants' differences in this work. Age significantly impacted both fixations and verbal responses. The older the participant, the higher the likelihood of a fixation in a hot spot ($p = 4.12e{-}08$) and the higher the likelihood of a global verbal response ($p = 0.01$). Pragmatic ability also predicted global verbal responding ($p = 9.919e{-}11$).

Secondly, the baseline condition predicted an increased likelihood of fixation in a hot spot ($p = 0.02$), yet the filtered condition predicted an increased likelihood of global verbal response ($p = 0.005$). Also, session order impacted fixations in hot spots ($p = 8.56e{-}08$) such that images seen in the second session,

regardless of condition, were far more likely to result in a fixation in a hot spot, however, no difference was found for global verbal behavior ($p = 0.19$).

Lastly, we review the results by image characteristics. These results are separated into semantic content of the image, luminance, and spatial frequency. The later 2 are the independent variables that were manipulated in the study as images were filtered or not filtered. At the semantic level of the image, we found a significant likelihood of a global verbal response ($p = 6.10e{-}07$) when an image contained a living object, however, no similar trend was found for eye fixations ($p = 0.82$). Image characteristics at the sensory perception level (i.e., luminance, spatial frequency) also played a role in global responding. We saw a significant likelihood of both a fixation in a hot spot and a global verbal response where the cold spot luminance approaches the ends of the luminance scale. Specifically, when the average luminance in the cold spot area is lower (i.e., darker) than the hot spots, there is a higher likelihood of a fixation in the hot spot, ($p = 0.007$). Whereas, when the average cold spot luminance average is higher (i.e., closer to white) the likelihood of giving a global verbal response is also statistically significant ($p = 0.003$). In other words, darker cold spots resulted in more fixations, yet lighter cold spots yielded more global verbal responses.

The analysis of spatial frequency did not yield significant findings for either hot spots or cold spots regarding fixations and verbal responding. However, because spatial frequency is dependent on luminance (Ellemberg et al. 1999), we then ran an interaction analysis where luminance and spatial frequency are combined into a new variable and compared to the fixation rate. At the interaction level of analysis, we see significant results for fixations in hot spots. The coefficient of hot spot spatial frequency was -0.15 and -0.16 for luminance hot spot, suggesting that the higher the luminance of a hot spot, the more powerful a lower spatial frequency becomes in predicting the likelihood of a fixation in a hot spot. This combination of luminance and spatial frequency provides a metric for the naturally occurring contrast created by terrain and textures—which adds complexity to the stimuli. Complex stimuli require more global visual processing which is known to be a challenge in autism (Bertone et al. 2005, p 128).

In summary we found several variables that improve the likelihood of global verbal responding. First, we found the older the participant, the greater the likelihood of fixations in hot spots and global verbal responses. Second, we found that more developed pragmatic skills based on the CELF-4 Social Pragmatic Score predicted more global verbal responses but did not impact the likelihood of a fixation in hot spots. Third, we found that the baseline condition (raw image) increased the likelihood of

Table 2 Results of the General Linear Regression models.

Variable/Model	Fixations in hot spots Est., std error, z value, Pr (>|z|)	Global verbal score Est., std error, z value, Pr (>|z|)
Age	0.44, 0.08, 5.48, 4.12e−08***	0.16, 0.06, 2.39, 0.01**
Social Pragmatic level	−0.20, 0.11, −1.83, 0.06	−0.69, 0.10, −6.48, 9.919e−11***
Condition	−0.077, 0.34, −2,24, 0.02*	0.87, 0.31, 2.75, 0.005***
Session order	0.81, 0.5, 5.35, 8.56e−08***	0.17, 0.13, 1.13, 0.19
Image order	0.00, 0.00, 1.84, 0.06	0.004, 0.004, 0.86, 0.38
Semantic level of image	0.04, 1.8, 0.22, 0.82	0.86, 0.17, 4.98, 6.10e−07***
Size of hot spots	0.15, 0.08, 1.72, 0.08	0.03, 0.08, 0.40, 0.68
Avg luminance in cold spots	−0.36, 0.13, −2.76, 0.007**	0.36, 0.12, 2.94, 0.003**
Avg luminance in hot spots	0.02, 0.10, 0.23, 0.81	−0.09, 0.09, −0.966, 0.33
Avg spatial freq of whole image	0.04, 0.11, 0.40, 0.68	−0.12, 0.10, −1.25, 0.21
Avg spatial freq of hot spot	−0.13, 0.07, −1.72, 0.08	−0.12, 0.07, 11.69, 0.08

Sig codes: 0 '***' 0.001, '**' 0.01 '*' 0.05.

fixation in a hot spot while the filtered condition improved the likelihood of global verbal response. Fourth, we found that trials in the second session were significantly more likely to yield fixations in hot spots—regardless of condition. Fifth, we found that darker, lower luminance in the cold spots improved likelihood of fixations in hot spots while higher, lighter luminance in cold spots increased the likelihood of global verbal responding. Lastly, we found the contrast or visual texture (luminance × spatial frequency) was most effective at producing fixations in hot spots when the hot spot was light and had lower spatial frequency. Next, we discuss the limitations of the study and implications of these findings.

Discussion

We hypothesized that filtering a scene to augment the socially salient features could serve as an assistive technology—thus empowering autistics children via access to global information in natural scenes. We found some evidence of this as well as other findings beyond our Research Questions. For example, we found a relation between global responding and age and pragmatic ability, we will discuss each significant variable. We begin with a discussion of the limitations to this work.

Study design (session order and condition). First, the language prompt given to participants could have been confusing. Although we provided training trials at the first session, the prompt "What is this picture about" was likely difficult for an autistic individual to answer, given the notorious difficulty this population demonstrates in formulating responses to Wh-Question forms. This is especially true for the younger children. A better prompt for future work could be "This picture is about…" after clear training on how to answer in a training phase. Secondly, the statistically significant session order finding warrants a deeper look into the reliability of eye gaze behaviors over time. Many studies of eye tracking in autism are conducted in a single session, thus possibly not revealing changes in fixations that occur over time. We intentionally spaced out our sessions to not expose participants to the same image in a single sitting of 10 min. This design choice resulted in significant differences in performance in the second session. Future work could explore the impact of the filter within a single session as well as consider running longer studies to understand the stability of eye gaze behavior over time and the impact of the filter.

Implications for assistive technology. There are several implications for the design of innovative assistive technologies based on image characteristics. Specifically, this work provides early evidence that global processing could be augmented by manipulating visual stimuli. It is well established that looking at key elements provides a common reference for communication (Clark 1996). Addressing sensory features to augment social communication is a novel approach to the design of assistive technologies such as a global filter.

Semantic content of image. It seems logical that global verbal responses would be more likely in picture for images that contained living aspects. One possibility beyond age and social pragmatic ability is more basic communication ability because living image labels may develop early or be more prominent in vocabulary. Further work could examine the role of vocabulary and age of acquisition in semantic representation. We found no difference in likelihood of a fixation in living or non-living images, which may be surprising as the canon of research on salience of social stimuli in autism demonstrates differences based on social and non-social stimuli (Frazier et al. 2017; Klin et al. 2002; Skwerer et al. 2019). This lack of impact based on semantic content on fixations supports our rationale for manipulating sensory-level characteristics to promote visual attention.

Luminance and spatial frequency. We found results for luminance and spatial frequencies ranges in areas of interest that are statically significant and therefore, should be considered along with other key variables such as age and language ability. This suggests a key role of sensory processing in the processing of global information that is commonly used to communicate about visual information.

Low-level complexity of image. We have examined the low-level complexity of an image and shown significant results of human factors, study design components and image characteristics. We designed a global filter to bring luminance to the mean to reduce the contrast between light and dark pixels; however, our findings suggest that dimming the brightness rather than averaging it may be more effective at directing eye gaze to hot spots. However, lighter cold spot areas were more predictive of verbal semantic representation. This contradictory finding has implications for the design of future filters and offers an opportunity to untangle the relationship between fixations and verbal responses.

In summary, we found significant results regarding the impact of luminance of the cold spots on both eye fixations and global responding—but in opposite directions.

Future work could explore a wider variety of eye tracking behavior for a broader perspective on visual attention such as saccades, blinks, and pupillometry (Sim and Bond 2021). Future work could also explore visual attention over time. For example, the integration that occurs over time could be supported by manipulating luminance and spatial frequency across the first moments of observation. A temporal adjustment to visual stimuli could potentially offer customized visual processing support for seeing and understanding the gist.

In summary, this work demonstrates the feasibility of priming semantic representations as well as shifting eye gaze to global areas by visually highlighting semantic regions of images. After evaluating several features of the study and the image characteristics, we found that low levels of luminance in the background of an image predicted higher rates of gazing at hot spots. This suggests images can be manipulated to support eye gaze to global regions of interest—the goal of which would be to see the main objects first thus reducing the processing overload of local details. This "preprocessing" of global areas allows for cognitive resources to be directed to areas that connote primary objects that can then be analyzed in more detail. Once improved, the filter could be automated for real time use on digital devices and eventually applied to 3D and real time spaces to assist in global processing of the physical world, see Fig. 2. Lastly, these findings could be considered for interventions for other neurodiverse communities with local interference such as Obsessive Compulsive Disorder (Yovel et al. 2005) or related patterns to global processing deficits such as Fragile X Syndrome, William's syndrome, dyslexia and dyspraxia (Grinter et al. 2010).

Visual attention is a complicated process "as information at different levels of the visual hierarchy is not equally likely to become conscious; rather, conscious percepts emerge preferentially at a global level" (Campana et al. 2016, p 5200). The first moment of viewing an image could be directed by a gaze prediction software that directs visual attention to primary areas of interest, and then an image could revert to its baseline state or lighter background to provide all the information contained within the image. Implications of this work could help build a

Fig. 2 Mockup of a Phone App that illustrates the implementation of the global filter for use by those wanting support with global salience. The phone interface shows a filtered image, a banner that states "global filter" to indicate the picture has been filtered and a text box that mimics social networking apps for chatting with friends. This figure is not covered by the Creative Commons Attribution 4.0 International License. Reproduced with permission of LouAnne E Boyd; copyright © LouAnne E Boyd, all rights reserved.

global filter to drive visual attention to global areas of images, video, and the real world[1,2,3,4].

Data availability
The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Notes
1 We build on the think-see-do concept introduced in (Brimmell et al. 2021) related to executive functioning, visual attention, and sports performance.
2 We borrow the term "semantic priming" from (Toro 2000) who examine automatic versus strategic semantic priming in Broca's aphasia.
3 This current work is part of an ongoing project that began a year earlier with a separate group of children with speech and language impairments and their SLPs to determine the type of filter to be used. Our pilot study with four different low-fidelity filters and baseline images (paper in submission). In that early work, the manually created filters revealed that blurring and desaturating the less relevant areas resulted in more global verbal responses. Also, prior to this study, we conducted the same low-fidelity probe with an autistic adult while collecting eye tracking data. The results informed the design of the current filter which automatically renders filtered images.
4 This current work is part of an ongoing project that began a year earlier with a separate group of children with speech and language impairments and their SLPs to determine the type of filter to be used. Our pilot study with four different low-fidelity filters and baseline images (paper in submission). In that early work, the manually created filters revealed that blurring and desaturating the less relevant areas resulted in more global verbal responses. Also, prior to this study, we conducted the same low-fidelity probe with an autistic adult while collecting eye tracking data. The results informed the design of the current filter which automatically renders filtered images.

References
Badcock DR, Clifford CWG, Khuu SK (2005) Interactions between luminance and contrast signals in global form detection. Vis Res 45:881–889. https://doi.org/10.1016/j.visres.2004.09.042

Baisa A, Mevorach C, Shalev L (2018) Can performance in Navon letters among people with autism be affected by saliency? Reexamination of the literature. Rev J Autism Dev Disord 6:1–12

Baisa A, Mevorach C, Shalev L (2020) Hierarchical processing in ASD is driven by exaggerated salience effects, not local bias. J Autism Dev Disord 51:666–676

Bertone A, Mottron L, Jelenic P, Faubert J (2005) Enhanced and diminished visuo-spatial information processing in autism depends on stimulus complexity. Brain 128:2430–2441. https://doi.org/10.1093/brain/awh561

Bouvet L, Rousset S, Valdois S, Donnadieu S (2011) Global precedence effect in audition and vision: evidence for similar cognitive styles across modalities. Acta Psychol138(2):329–35

Brimmell J, Edwards EJ, Smith M, Vaughan RS (2021) Think, see, do: executive function, visual attention, and soccer penalty performance. Sport Exerc Perform Psychol 10:290–309

Burton GJ, Moorhead IR (1987) Color and spatial structure in natural scenes Appl Optics 26(1):157–70

Campana F, Rebollo I, Urai AE, Wyart V, Tallon-Baudry C (2016) Conscious vision proceeds from global to local content in goal-directed tasks and spontaneous vision. J Neurosci 36:5200–5213

Chawarska K, Shic F (2009) Looking but not seeing: Atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. J Autism Dev Disord 39:1663–1672

Clark HH (1996) Using language. Cambridge University Press, Cambridge

Courchesne E, Pierce K (2005) Why the frontal cortex in autism might be talking only to itself: local over-connectivity but long-distance disconnection. Curr Opin Neurobiol 15:225–230. https://doi.org/10.1016/j.conb.2005.03.001

de Jong MC, vanEngeland H, Kemner C (2008) Attentional effects of gaze shifts are influenced by emotion and spatial frequency, but not in autism. J Am Acad Child Adolesc Psychiatry 47(4):443–454

Deruelle C, Rondan C, Gepner B, Tardif C (2004) Spatial frequency and face processing in children with autism and asperger syndrome. J Autism Dev Disord 34:199–210

Dickson CA, Wang SS, Lombard KM, Dube WV (2006) Overselective stimulus control in residential school students with intellectual disabilities. Res Dev Disabil 27(6):618–31

Doehring P (2021) Does the arc of science bend towards impact? Four decades of empirical research published in JADD since the DSM-III. J Autism Dev Disord 51:4411–4421

Dube WV et al. (2010) Observing behavior and atypically restricted stimulus control. J Exp Anal Behav94(3):297–313

Ellemberg D, Lewis TL, Liu CH, Maurer D (1999) Development of spatial and temporal vision during childhood. Vis Res 39:2325–2333

Eskicioglu A, Fisher P (1995) Image quality measures and their performance. IEEE Trans Commun 43:2959–2965. https://doi.org/10.1109/26.477498

Fan J, Posner M (2004) Human attentional networks. Psychiatr Praxis 31:210–214

Field DJ (1987) Relations between the statistics of natural images and the response properties of cortical cells. JOSA A 4:2379–2394. https://doi.org/10.1364/JOSAA.4.002379

Frazier TW et al. (2017) A meta-analysis of gaze differences to social and nonsocial information between individuals with and without autism. J Am Acad Child Adolesc Psychiatry 56(7):546–555

Gargaro BA et al. (2018) Attentional mechanisms in autism, ADHD, and autism-ADHD using a local–global paradigm. J Atten Disord 22:1320–1332

Grinter EJ, Maybery MT, Badcock DR (2010) Vision in developmental disorders: Is there a dorsal stream deficit? Brain Res Bull 82:147–160

Grinter EJ et al. (2009) Global visual processing and self-rated autistic-like traits. J Autism Dev Disord 39:1278–1290

Gross TF (2005) Global–local precedence in the perception of facial age and emotional expression by children with autism and other developmental disabilities. J Autism Dev Disord 35:773–785

Guy J, Mottron L, Berthiaume C, Bertone A (2016) A developmental perspective of global and local visual perception in autism spectrum disorder. J Autism Dev Disord 49:2706–2720

Happé F, Booth R, Charlton RA, Hughes C (2006) Executive function deficits in autism spectrum disorders and attention-deficit/hyperactivity disorder: examining profiles across domains and ages. Brain Cogn 61:25–39

Hughes HC, Nozawa G, Kitterle FL (1996) Global precedence, spatial frequency channels, and the statistics of natural images. J Cogn Neurosci 8:197–230

Jolliffe T, Baron-Cohen S (1999) A test of central coherence theory: linguistic processing in high-functioning adults with autism or asperger syndrome: Is local coherence impaired? Cognition 71:149–185

Kaliukhovich DA et al. (2020) Visual preference for biological motion in children and adults with autism spectrum disorder: an eye-tracking study. J Autism Dev Disord 51:2369–2380

Katagiri M, Kasai T, Kamio Y, Murohashi H (2013) Individuals with asperger's disorder exhibit difficulty in switching attention from a local level to a global level. J Autism Dev Disord 43:395–403

Kihara K, Takeda Y (2019) The role of low-spatial frequency components in the processing of deceptive faces: a study using artificial face models. Front Psychol 10:1468. https://doi.org/10.3389/fpsyg.2019.01468

Klin A, Jones WR, Schultz RT, Volkmar FR, Cohen DJ (2002) Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. Arch Gen Psychiatry 59(9):809–16

Koegel LK (2000) Interventions to facilitate communication in autism. J Autism Dev Disord 30:383–391

Koegel RL, Wilhelm H (1973) Selective responding to the components of multiple visual cues by autistic children. J Exp Child Psychol 15:442–453. https://doi.org/10.1016/0022-0965(73)90094-5

Koh HC, Milne E, Dobkins KR (2010) Spatial contrast sensitivity in adolescents with autism spectrum disorders. J Autism Dev Disord 40:978–987

Koldewyn K, Jiang Y, Weigelt S, Kanwisher N (2013) Global/local processing in autism: not a disability, but a disinclination. J Autism Dev Disord 43:2329–2340. https://doi.org/10.1007/s10803-013-1777-z

Liechty J, Pieters R, Wedel M (2003) Global and local covert visual attention: evidence from a Bayesian hidden Markov model. Psychometrika 68:519–541. https://doi.org/10.1007/BF02295608

Pasternak T, Merigan WH (1981) The luminance dependence of spatial vision in the cat. Vis Res 21:1333–1339. https://doi.org/10.1016/0042-6989(81)90240-6

Skwerer DP et al. (2019) Do minimally verbal and verbally fluent individuals with autism spectrum disorder differ in their viewing patterns of dynamic social scenes? Autism 23:2131–2144

Ploog BO (2010) Stimulus overselectivity four decades later: a review of the literature and its implications for current research in autism spectrum disorder. J Autism Dev Disord 40:1332–1349

Richardson DC, Dale R (2005) Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. Cogn Sci 29:1045–1060. https://doi.org/10.1207/s15516709cog0000_29

Rieth SR, Stahmer AC, Suhrheinrich J, Schreibman L (2015) Examination of the prevalence of stimulus overselectivity in children with ASD. J Appl Behav Anal 48:71–84. https://doi.org/10.1002/jaba.165.

Rinehart NJ, Bradshaw JL, Moss SA, Brereton AV, Tonge BJ (2000) Atypical interference of local detail on global processing in high- functioning autism and asperger's disorder. J Child Psychol Psychiatry Allied Discip 41:769–778

Robertson CE, Baron-Cohen S (2017) Sensory perception in autism. Nat Rev Neurosci 18:671–684. https://doi.org/10.1038/nrn.2017.112

Sean V (2020) Connecting the dots for people with autism: a data-driven approach to designing and evaluating a global filter. Ph.D. thesis. Chapman University, California

Shulman GL, Sullivan MA, Gish K, Sakoda WJ (1986) The role of spatial-frequency channels in the perception of local and global structure. Perception 15:259–273. https://doi.org/10.1068/p150259

Shulman GL, Wilson J (1987) Spatial frequency and selective attention to local and global information. Perception 16:89–101. https://doi.org/10.1068/p160089

Sim G, Bond R (2021) Eye tracking in child computer interaction: challenges and opportunities. Int J Child-Computer Interact. https://doi.org/10.1016/j.ijcci.2021.100345

Song Y, Hakoda Y (2015) Lack of global precedence and global-to-local interference without local processing deficit: a robust finding in children with attention-deficit/hyperactivity disorder under different visual angles of the Navon task. Neuropsychology 29:888–894. https://doi.org/10.1037/neu0000213

Spering M, Carrasco M (2015) Acting without seeing: eye movements reveal visual processing without awareness. Trends Neurosci 38:247–258

Toet A (2006) Gaze directed displays as an enabling technology for attention aware systems. Comput Hum Behav 22:615–647. https://doi.org/10.1016/j.chb.2005.12.010

Toro JFD (2000) An examination of automatic versus strategic semantic priming effects in Broca's aphasia. Aphasiology 14:925–947. https://doi.org/10.1080/02687030050127720

Vision NRCUCo (1985) Two modes of visual processing. National Academies Press, Washington

Wang S et al. (2015) Atypical visual saliency in autism spectrum disorder quantified through model-based eye tracking. Neuron 88:604–616. https://doi.org/10.1016/j.neuron.2015.09.042

Wedel M, Pieters R, Liechty JC (2008) Attention switching during scene perception: how goals influence the time course of eye movements across advertisements. J Exp Psychol Appl 14(2):129–38

Xu J, Jiang M, Wang S, Kankanhalli MS, Zhao Q (2014) Predicting human gaze beyond pixels. J Vis 14:28–28. https://doi.org/10.1167/14.1.28

Yanulevskaya V, Uijlings JRR, Geusebroek JM, Sebe N, Smeulders AWM (2013) A proto-object-based computational model for visual saliency. J Vis 13(13):27

Yovel I, Revelle W, Mineka S (2005) Who sees trees before forest? The obsessive-compulsive style of visual attention. Psychol Sci 16:123–129. https://doi.org/10.1111/j.0956-7976.2005.00792.x

Competing interests

The authors declare no competing interests.

Ethical approval

This study was approved by the University's ethical committee IRB# 19-167. All the study was performed in accordance with the Declaration of Helsinki.

Informed consent

All the legal guardians of the children participating in this study signed an informed consent before conducting the study. Before starting each session, each participant provided consent to do the study.

Additional information

Correspondence and requests for materials should be addressed to LouAnne Boyd.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.