UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Development of a Computing App for the Evaluation of Executive Functions in Autism Spectrum Disorder's patients

Anne-Teresa Martins Pesnot

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em: Engenharia Clínica e Instrumentação Médica

Dissertação orientada por:
Cláudia Bandeira de Lima
Hugo Alexandre Ferreira

2019

ACKNOWLEDGMENTS

The first person I want to thank is the most important person in my life: MÃE. I would’ve never been able to achieve anything without your love, trust and unconditional support through everything in our lives. Thanks mammy, we will always have each other.

A big thank you to Zé for all the patience, kindness and for making me see what its like to have a real father.

I don’t think I will ever have enough words to thank my boyfriend, João. This work is also yours and thank you so much for all the love and support, even when I didn’t deserve it. You are the most amazing person, my best friend and you have been my rock every day and I don’t want that to ever change.

I also want to thank my amazing family for all the support through this journey, and especially my cousin Nuno for all the late nights and constant motivation.

To Professor Hugo Ferreira, a big thank you. I could never do it without your guidance, advice, patience and kindness. It was a pleasure working with you.

To my co-advisor, Dra. Cláudia Bandeira de Lima, thank you so much for everything. I feel lucky to have developed this project with you and it was such an amazing experience to see your work and learn from you.

I also want to thank all the parents and children that participated in this study for your kindness and availability in such a short time. Thank you as well to the institutions that helped develop and test this work. To Centro Hospitalar Universitário Lisboa Norte and Centro de Neurodesenvolvimento LógicaMentes, a big acknowledgment for providing the means to prove the potential of this work.

To my best friend Rafaela, a major thanks. Even far away you never let me give up and always encouraged me to believe in myself and push through.

Finally, I want to thank my friends for all their support, friendship and for making me do this even when I didn’t think it was possible. You are the best friends anyone can have. A special thank you to Vítor for all your help and expertise.

ABSTRACT

Autism Spectrum Disorders (ASD) is a neurodevelopmental disorder that show themselves early in a person's life and persist throughout their lifespan, having diverse consequences in social life and even in cognitive development. The diagnosis of ASD is complicated and involves a series of tests with the patient and their caretaker to identify warning signs and assessing the severity of the disorder.

Previous studies have suggested that eye movements are a reflection of cognitive functions that are presented differently in cases of ASD. Using eye tracking devices and software, eye movements can be analysed to evaluate the executive profile in ASD patients and help with a possible intervention.

In this dissertation, a game that consists of two tasks (Coding and Puzzles) typically used to evaluate ASD in children aged 6 to 10 years was developed and integrated with a low cost eye tracker.

Results suggest that ASD children have more difficulty with tasks associated with coding and present eye movements that are substantially different than those of children with a typical development.

Based on these results, further work can be done in order to develop a robust tool that integrates the presently used methods in digital format with eye tracking hardware and software and facilitates the evaluation of and intervention in ASD in clinical psychology.

Keywords: Autism; Eye tracking; Evaluation; Medical software; Computer application.

RESUMO

A Perturbação do Espectro do Autismo (PEA) é uma perturbação do neurodesenvolvimento que se manifesta cedo na vida de uma pessoa e persiste ao longo da sua vida, tendo diversas consequências a nível social, nas atividades do quotidiano e até a nível do desenvolvimento cognitivo em alguns casos. A PEA afeta não só o indivíduo diagnosticado com a doença, mas também toda a dinâmica familiar. O diagnóstico de PEA pode ser bastante complicado e envolve múltiplos questionários e testes com o doente e os familiares mais próximos de modo a identificar os sinais de alarme e avaliar a gravidade da perturbação.

Diversos estudos prévios sugerem que o olhar (movimentos oculares) pode refletir funções cognitivas que se encontram alterada nos casos de PEA. Através da utilização de equipamentos de rastreamento de olhar e software especializado, é possível analisar os movimentos oculares e avaliar o perfil executivo em doentes com PEA. Em alguns casos, o desenvolvimento de software especializado é, também, útil na intervenção da doença.

Estudos anteriores demonstraram diferenças significativas a nível de execução das tarefas da Wechsler Intelligence Scale for Children – III (WISC – III) entre crianças com PEA e crianças com outras perturbações do neurodesenvolvimento. Assim, o objetivo do projeto desenvolvido no âmbito desta dissertação é verificar se existem diferenças significativas a nível de execução entre crianças com PEA e crianças com desenvolvimento típico (DT).

Este projeto consiste num jogo com duas tarefas normalmente utilizadas (tarefas adaptadas da WISC– III) no ambiente clínico na avaliação de PEA em crianças com idades compreendidas entre os 6 e os 10 anos de idade.

A primeira tarefa trata-se de uma tarefa de código com o objetivo de avaliar a atenção, processamento sequencial, memória a curto prazo, capacidades psicomotoras, velocidade de execução e planeamento da tarefa. Durante esta tarefa os participantes têm como objetivo associar cinco imagens representativas de três temáticas diferentes (formas geométricas, animais da quinta e meios de transporte) à letra correta, seguindo um código que está sempre visível no topo do ecrã. A tarefa tem três níveis (cada um com uma temática específica) e em cada nível o número de associações necessárias até completar o objetivo aumenta, de modo a aumentar também a dificuldade. As imagens são apresentadas no lado direito do ecrã numa caixa designada por "Inventário" e as associações devem ser feitas da esquerda para a direita ao longo das linhas, pela ordem apresentada, de acordo com o código no topo do ecrã.

A segunda tarefa incluída neste jogo é a composição de objetos. Neste caso pretende-se avaliar a organização visual e espacial, memória visual, noção do todo a partir das suas partes e coordenação visual e motora. Nesta tarefa os participantes têm como objetivo observar uma imagem completa durante 30 segundos e, de seguida, reconstruir essa imagem a partir de um número de peças que aumenta ao longo de três níveis, juntamente com a complexidade da imagem mostrada inicialmente. As peças que compõem a imagem são apresentadas do lado direito do ecrã, numa caixa de "Inventário" e os participantes são livres para pegar em qualquer uma das peças e tentar colocá-la na posição correta.

A tarefa de código foi escolhida por ser a tarefa que apresentou mais diferenças entre os grupos nos estudos feitos anteriormente. Por outro lado, a tarefa de composição de objetos foi incluída como tarefa de validação e também com o intuito de comparar se existem diferenças entre estas duas populações e as populações estudadas anteriormente (doentes com outras doenças do neurodesenvolvimento).

O jogo foi integrado com um equipamento de rastreamento de olhar de baixo custo de modo a recolher dados sobre os movimentos oculares enquanto as crianças realizavam as tarefas. O processo de calibração do equipamento foi único para cada um dos participantes (dos quais 14 eram crianças com DT e 8 eram crianças diagnosticadas com PEA) e foi baseado na fixação do olhar em pontos específicos do ecrã, conforme as indicações fornecidas passo a passo pelo software do sensor de rastreamento do olhar.

Os resultados obtidos relativos à primeira tarefa sugerem que as crianças com PEA apresentam maiores dificuldades nas tarefas de código, verificando-se maior número de erros e um tempo total de execução da tarefa aumentado para as crianças com PEA. A justificação destes resultados prende-se também com a dificuldade cognitiva das crianças com PEA relativas às áreas avaliadas.

Também o padrão de movimentos oculares destas crianças é substancialmente diferente quando comparado com crianças com DT. Os participantes com PEA aparentam ter sentido uma maior necessidade de consultar o código durante a tarefa de código e continuaram a apoiar-se nesta consulta durante os três níveis da tarefa; por outro lado as crianças de DT mostraram uma necessidade muito menor de consultar o código e, à medida que avançavam no jogo, consultaram-no cada vez menos. Além disto, os padrões de movimentos oculares dos participantes com PEA sugerem uma maior aptidão por certas letras do código enquanto os padrões dos participantes com DT sugerem um olhar mais organizado e não concentrado em letras específicas.

Relativamente à segunda tarefa, as diferenças entre os dois grupos de participantes não foram tão acentuadas, existindo uma progressão no número de erros por nível quase linear e semelhante entre os dois grupos; no entanto, o tempo total de execução da tarefa no nível 3 aumentou substancialmente no caso dos participantes com PEA, o que mais uma vez sugere dificuldade em lidar com uma maior quantidade de estímulos e poderá indicar maior dificuldade em manter a integridade da imagem enquanto todo quando a desfragmentação é maior. Não obstante, a percentagem de participantes com PEA que completou a tarefa de composição foi muito elevada e superior à da primeira tarefa, o que sugere uma preferência destas crianças pela tarefa de composição, como era esperado de acordo com a literatura acerca do autismo e das diferenças cognitivas e sociais associadas a esta perturbação do neurodesenvolvimento.

Nesta segunda tarefa também se verificaram diferenças nos padrões de movimentos oculares: os participantes com PEA focaram-se maioritariamente num único detalhe em cada imagem enquanto os participantes com DT tentaram focar vários pontos de cada imagem de modo a reter mais detalhes da mesma. Além disto, alguns participantes com PEA percorreram com o olhar o rebordo da primeira imagem de forma vaga, indicando uma possível falta de concentração na tarefa proposta.

Os questionários realizados a cada participante após a sua participação no estudo revelaram que a grande maioria das crianças (95,8%) gostou do jogo e 73,1% de todos os participantes conseguiram terminar o jogo na sua totalidade. Apesar disto, é de notar que nem todas as crianças com PEA tiveram uma reação positiva ao jogo e precisaram de muito encorajamento e motivação para completar as tarefas que lhes eram propostas, enquanto as crianças com DT mostraram todas uma reação positiva ao jogo e completaram-no sem grande necessidade de encorajamento por parte dos pais/encarregados de educação ou dos investigadores presentes durante os testes.

Com base nestes resultados, conclui-se que existem diferenças claras e fáceis de identificar nos padrões de movimento ocular de crianças com DT e crianças com PEA. Estas diferenças podem vir a ser utilizadas em trabalho a desenvolver no futuro de modo a obter uma ferramenta robusta, que integre os métodos utilizados em ambiente clínico – que foram neste projeto adaptados para formato digital – com hardware e software de rastreamento de olhar – e que facilite a avaliação do perfil executivo ou a possível intervenção nas PEA em psicologia clínica.

Palavras – chave: Autismo; Rastreamento do olhar; Avaliação; Software médico; Aplicação para computador

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

ASD Autism Spectrum Disorder

DSM – 5 Diagnostic and Statistical Manual of Mental Disorders

E Number of Errors

GDPR General Data Protection Regulation

GPL General Public License

IDD Intellectual Developmental Disorder

IR Infrared

LD Language Disorder

LT Latency Time

ND Neurodevelopmental Disorder

OND Other Neurodevelopment Disorders

PCCR Pupil Centre Corneal Reflection

SDK Software Development Kit

std Standard Deviation

TD Typical Development

TT Total Time of Execution

UI User Interface

WISC-III Wechsler Intelligence Scale for Children – III

1 INTRODUCTION

1.1 TECHNOLOGY AND HEALTH – BIOMEDICAL ENGINEERING

Technology and innovation have become an integral part of modern life in industrialized nations. The area of medicine and healthcare service is not an exception to these advancements, in fact, it is one of the areas where they are noticed the most. The practice of medicine has a long history, ranging from the prehistoric era through various ancient societies, like Ancient Egypt, where the first medicinal texts came from, all the way to modern medicine.

Despite this long history and the development of the practice of medicine throughout the ages, the development of a health care system based on advanced technology and capable of a wide range of effective diagnostic and therapeutic procedures is a relatively recent accomplishment. Because of this advancement in technology in recent decades, engineering professionals have become more and more involved in medicine and healthcare. This integration of technology and medicine gave rise to the discipline of biomedical engineering [1].

The importance of the biomedical engineer becomes clear when we consider that many of the challenges faced by healthcare professionals nowadays involve aspects of device and systems analysis, design and practical applications – these aspects are fundamental to engineering practice. When specific healthcare related problems require the application of principles known to any of the engineering disciplines, it is the biomedical engineer that bridges the gap between those disciplines and the professionals facing those problems. In addition to that, advancements in healthcare lead to an increase in life expectancy, which in turn leads to an increasing demand for biomedical devices and procedures such as hip and knee replacements (for the older population), and as society becomes more and more aware of new developments in technology – because of the incredible amount of information that is so readily available in current times –increasing numbers of people will seek biomedical solutions to their health problems, making biomedical engineering a rapidly growing field whose relevance is very clear, both today and in the near future [2].

It is noteworthy, however, that the breadth of activity of biomedical engineers has increased significantly since being primarily concerned, in the 1950s and 1960s, with the development of devices such as those mentioned in the previous paragraph; nowadays, biomedical engineering encompasses areas as diverse as the detection, measurement and monitoring of physiologic data, the computer analysis of patient-related data and the so called rehabilitation engineering, which is connected to the development of therapeutic and rehabilitation procedures and devices.

One example of physiologic data which is interesting to measure and analyse is eye movement data, especially in cases of psychological or cognitive issues, because this data can be studied in order to compare the way "healthy" people move their eyes when performing visual tasks with the way an "unhealthy" person moves theirs during the same task. This data can be captured relatively easily using eye tracking devices.

1.2 CONTEXT

This dissertation, written by Anne-Teresa Martins Pesnot, student of the integrated master's degree in Biomedical Engineering and Biophysics in Faculdade de Ciências da Universidade de Lisboa, was co-supervised by Cláudia Bandeira de Lima, PhD, a clinical psychologist from Hospital de Santa Maria, Lisbon. The internal supervisor was Hugo Ferreira, PhD, professor at the Faculdade de Ciências da Universidade de Lisboa.

In this project the goal was to evaluate and compare the performance and the eye movement data of children diagnosed with autism and typical development children in various visual reasoning tasks, using a computer application designed specifically to mimic various tests used in the process of diagnosing the comorbidity of Autism and laying the groundwork for future attempts at intervention and possibly therapy in some selected cases.

The selected device for the task proposed was Tobii 4C eye tracker, which is a low-cost commercial device (screen based), used in gaming and research and has a cost of 150 – 170€. Although there are some limitations in the use of this device, if it is proven this can be used for this application, it could be possible to reduce the costs of intervention in Autism Spectrum Disorders (ASD) and make it more accessible to the final user.

1.3 STRUCTURE OF THE DISSERTATION

The present dissertation is constituted by six chapters, where chapter 1 is the present introduction.

In chapter 2, various theoretical concepts are thoroughly explained to integrate the reader within the context of this dissertation. It features a revision about eye tracking, ASD and a state-of-the-art about eye tracking in ASD and its various applications. This chapter was included in order to enlighten the reader of how the eye tracking technology can be useful to study and intervene in ASD.

Chapter 3 contains a description of the materials and methods used throughout this dissertation. It includes the description of the project implementation and app development. In this chapter is also described the app testing process (participant description and procedure). Finally, it also explains the measures obtained from the testing phase.

Chapter 4 exhibits all the results obtained from the methods described in Chapter 3 and Chapter 5 presents a discussion over the results shown.

Finally, in chapter 6 the main conclusions drawn from the overall results obtained in this dissertation are presented and discussed, as well as limitations of this work and future steps.

2 BACKGROUND

2.1 EYE TRACKING

Eye tracking is the process of measuring the motion of the eyes relative to the head or the point of gaze (i.e. the focal point or where the subject is looking).

There are different types of eye trackers, and the technology they use vary with their type. Eye-attached tracking is a methodology that makes use of an attachment like a contact lens with an embedded mirror or magnetic field sensor, and what is measured is the movement of this attachment relative to a fixed point, with the assumption that the lens does not shift significantly during the eye movement. It is also possible to track eye movement using electric potentials measured during an electrooculogram. These, however, are invasive methods and thus are not recommended for studies like this one which involve subjects with severe psychological issues [3].

The non-invasive alternative and the technique used by Tobii© in their screen-based eye trackers is pupil centre corneal reflection (PCCR). In this technique an infrared light source is used to illuminate the eye causing highly visible reflections and then a camera captures an image of the eye showing these reflections (see Figure 2.1). This image is then processed to identify the reflection of the light source on the cornea and in the pupil. After identifying these reflections, a vector formed by the angle between them can be calculated and the direction of this vector, combined with other geometrical features, is then used to calculate the gaze point [4].

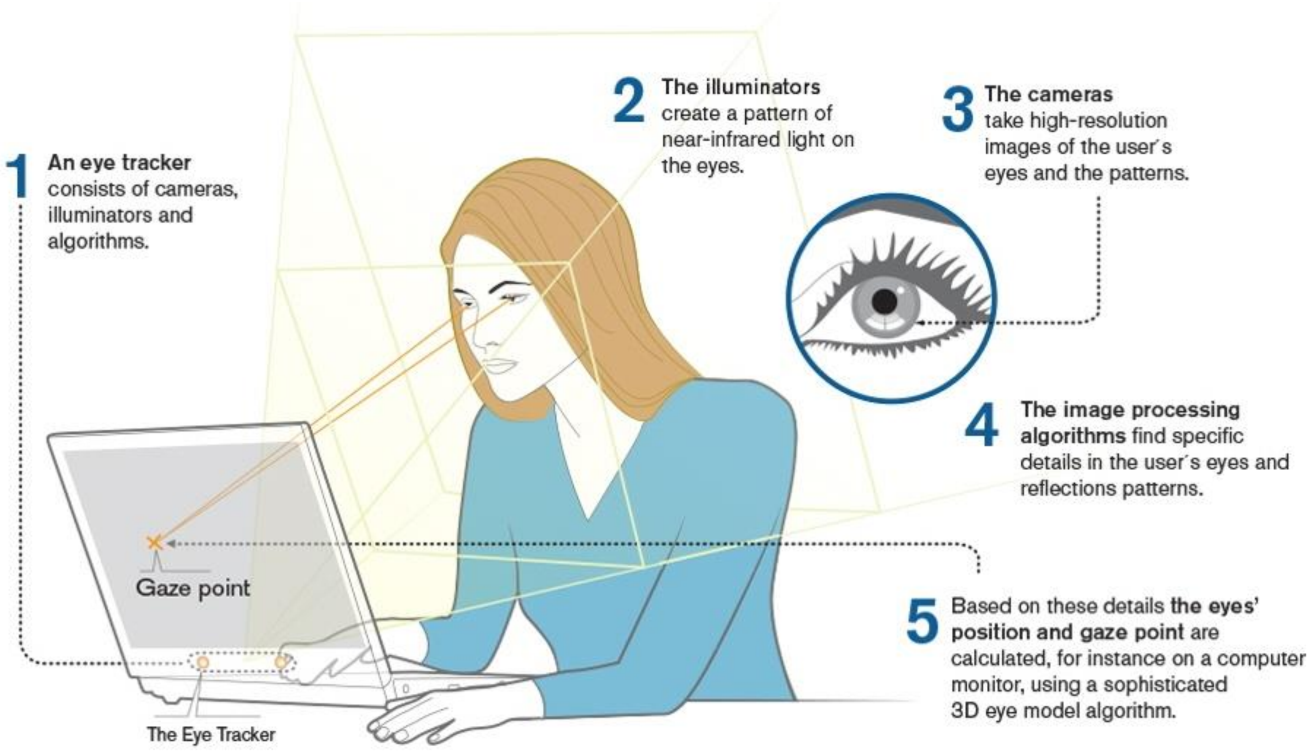


Figure 2.1: Functioning of screen-based eye tracker. [Figure adapted from [4]]

2.1.1 Calibration

Before eye movement can be acquired however, the eye tracking device must be calibrated. Calibration is the process through which the geometrical features of a subject's eyes are estimated in order to calculate the gaze point as accurately as possible for each subject.

During the calibration procedure the user is asked to look at specific points on the screen, also known as calibration dots. Meanwhile the eye tracker collects several images of the user's eyes and analyses them, integrating the results in an internal 3D model of an eye and calculating the gaze point for each image sample. The calibration procedure, like the eye tracking procedure in general, is only negatively impacted by head movement if the subject moves too fast, turns his head to the side or moves to the edges of the trackbox (the field-of-view of the sensor), because the eye tracker possesses two different cameras, which means there are always two different sources of information regarding the position of the eyes and head.

This double input setup contributes to the accuracy and precision of the system, which are important concepts to understand in order to know how one can evaluate the recorded data (see Figure 2.2) [5]. Accuracy is defined as the average difference between the true gaze position and the recorded position, whereas precision measures the variation of the recorded data via the Root Mean Square of successive samples. Accuracy is especially important because it varies across subjects and experimental conditions, because it depends on specific subject properties, illumination in the test environment, stimuli properties, calibration quality, data collection procedure and the eyes' position in the trackbox [6], [7].

Good precision, poor accuracy

Good accuracy, poor precision

Good accuracy, good precision

Poor accuracy, poor precision

Figure 2.2: Difference between accuracy and precision. [Figure adapted from [5]]

2.1.2 Heatmaps

A good way to visualise and analyse the gaze point data collected using an eye tracking device is using a heatmap (Figure 2.3) [8]. A heatmap is a graphical representation of data stored in a matrix (in this case, x and y coordinates and timestamps) using a system of color-coding to represent different values.

These representations are used in various forms of analytics but are most commonly used in order to show user behaviour in a format that is easy to read. They are used in various websites to show where the user has clicked, how far down the user has scrolled or, as is the case in this study, to display the results of eye tracking tests. Heatmaps are a very visual tool of analysis, which makes them ideal to analyse data that relates to positions on a plane over a period of time by making the analysis intuitive and accessible at first glance [9].

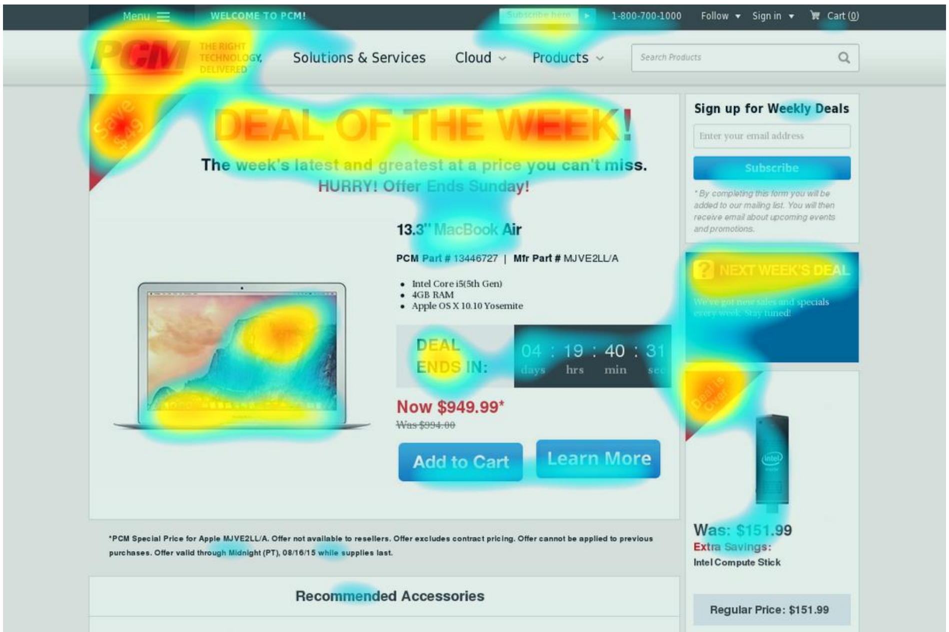


Figure 2.3: Example of an eye tracking heatmap. [Figure adapted from [8]]

2.1.3 Eye movements

The characteristics of the human eye limit the way we extract visual information from the world around us. Visual acuity, for instance, decreases rapidly as we move farther away from the centre of our field of view. To compensate for this decrease in visual acuity, our eyes are capable of performing both ample and minute movements in order to keep the object of interest in focus and in the centre of our vision.

When analysing a static object while our heads are relatively still, our eyes perform mainly saccades and fixational movements.

Saccades are a type of movement that is used to move the fovea rapidly from one point of interest to another. During a saccade, which can be triggered voluntarily or involuntarily, both of our eyes move in the same direction. The latency time of a saccade is dependent on the task and the duration of the movement itself is also variable, being linearly correlated with its amplitude (longer saccades take more time to execute, because the eye moves at a limited speed that depends on the subject). Curiously, the end point of a saccade cannot be changed once the movement has started; when the eye starts moving it already has a “destination” and it cannot be changed while the movement is taking place, which doesn’t happen with other movements and muscles like for example, an arm movement. Despite this the trajectory that the eye follows until it reaches the end point is not necessarily linear and can vary greatly [10].

Fixations, on the other hand, can be defined as the times when our eyes stop scanning the whole scene and hold the central foveal vision in place so that the visual system can take in detailed information about what we are looking at. A fixation is composed of slower and minute movements (microsaccades, tremor and drift) that help the eye align with the target and minimize perceptual fading. The duration of a fixation varies with the individual and with the task that is being performed; tasks that require the acquisition of more visual information generally present longer fixation times.

In more dynamic situations, where either we are moving or the object is, we want to look at is moving, or both, other eye movements are triggered in order to keep the fovea aligned with the object of interest. To help us focus on objects that are at different distances, we use vergence movements, which are generally slower than saccades and in which the left and right eye move in opposite directions (converging or diverging, depending on whether the object is near or far, respectively). To keep a moving object in focus we use the smooth pursuit movement which keeps the fovea aligned with the object as it moves and cannot be triggered voluntarily in the absence of a moving target; the eye moves at a maximum speed of about 30 degrees/second (although some individuals can reportedly smooth pursuit at velocities as high as 100 degrees/second) and when the target’s speed is higher than this value, the eye simply performs saccades to catch up with the target as it moves away from the centre of the field of view. Lastly, to keep an object in focus as we ourselves are moving or our head is moving (or both) we perform what is called the vestibular ocular reflex, in which our eyes move in the opposite direction of the head at the same speed in order to keep the fovea aligned with the object.

From an eye tracking testing perspective, the more interesting types of eye movements are the saccades and the fixations, because most of these tests take place while at rest and looking at a fixed plane. But fixations are not as simple to detect and record as one might think. Fixations, from an eye tracker data analyst’s perspective, are built from the basic output of the eye tracking system, the gaze points. These gaze points are instantaneous spatial locations and as such have an (x, y) coordinate and a timestamp corresponding to their measurement. Fixations can then be thought of as a collection of gaze points that share an (x, y) coordinate over a certain duration (normally between 50 ms and 600 ms).

It is noteworthy however that fixations are not real in the sense that they cannot be directly measured. Instead they are constructions, mathematical outputs that result from putting the raw gaze point data through what can be called a “fixation filter” that selects gaze points that can be converted to a fixation point using a conversion algorithm; this proved to be a limitation in this study, because access to the Tobii Pro software packages was not facilitated and the fixation algorithm is not available in the free software version.

Fixations are nevertheless noteworthy because they reveal useful information about attention, visibility, mental processing and understanding, because eye movement can be considered an executive function – part of the human body’s set of cognitive functions.

2.2 VISION AND EXECUTIVE FUNCTION

Cognitive functions are, broadly speaking, all means and mechanisms of acquiring information or, in other words, the functions that we as humans perform that ultimately lead to knowledge. They encompass logical reasoning, memory, attention and language [11].

Executive functions differ from cognitive functions in the sense that they do not lead to an acquisition of knowledge. They are, however, integral part of our ability to interact with the world based on our knowledge.

A universally accepted definition for executive function does not exist, but it is possible to identify them as complex, higher order processes moderated by the frontal lobe of the human brain. Attention (the ability to direct cognitive resources towards certain aspects of the environment or certain tasks), working memory (recalling and using important information for the execution of a task up until the execution itself) and motor planning (the act of planning a movement or set of actions with a specific goal in mind) are generally considered executive functions [12].

As was previously mentioned, eye movements serve the purpose of keeping the fovea aligned with the target; because of that, they determine what information reaches our higher cortical centres and are therefore critically important for vision, attention, and memory. Since they determine what we see, they are also critical for the aforementioned executive functions.

A great number of studies has been done investigating voluntary eye movement performance in children and adults diagnosed with autism. These studies have shown that executive function abilities are delayed in autistic patients; their voluntary abilities seem to develop at a slower rate and to a lower level. These studies also suggest that the executive function deficits in autistic patients critically involve the prefrontal cortex [13], [14].

Various studies have shown that eye movements reflect attention shifts, visual perception and the underlying organization of mental representations when performing a visual search of a scene [15]. Furthermore, longer and more frequent fixations as opposed to shorter saccades are an indicator of higher difficulty of understanding the visual scene or text [16], [17].

Other studies performed on linguistic problems such as dyslexia have also shown poorer stability in fixations, difficulties in smooth pursuit movements, lower amplitude vergence movements and shorter saccades in dyslexic children [18], which further corroborates the connection between eye movements and higher cognitive functions such as memory, problem solving, decision making, language and attention [4].

2.3 STATE-OF-THE-ART

Studies applying eye tracking technology to research related to ASD and possible methods of diagnosing them have shown differences in fixation times and fixation areas, but these studies have mostly been done using socially salient stimuli such as facial expressions and short video clips where people talk either to each other or to the audience [19], [20].

Researchers at Osaka University have, in 2010, conducted a study using eye tracking and short video clips of popular children's shows with the objective of trying to develop a quantitative scale for identifying individuals with autism based on gaze measurement that could be applied to both children and adults and thus facilitate the diagnosis of autism [4].

These researchers found, using screen-based eye tracking technology, that typical development individuals tended to focus on the speaker's face or the centre of the screen, reflecting standard gaze behaviour, while individuals with autism tended to look away from the speaker and focus on the periphery of the screen, as well as showing a preference to look at letters instead of other points of the image being shown [21].

In more recent years there have been numerous studies involving eye tracking and autism. In 2017 Michael Murias et al. performed a study with screen-based eye tracking and 25 children diagnosed with ASD involving a viewing of a videotape of an actor speaking directly to them and their results suggest the validity of using eye-gaze measurements as a biomarker associated with social communication abilities in children with ASD [22].

In 2018 Adrienne Moore et al. performed a study in which 227 toddlers (76 of which were diagnosed with ASD) watched a 90s video of dynamic geometric images paired with social images of children interacting and found that ASD toddlers spent much more time fixating on the geometric images when compared with other groups. In this study the objective was to identify a consistent, early-onset indicator of a subtype of ASD called GeoPref, characterized by a clear preference for geometric shapes over other types of images; the viewing of the 90s video – called the Complex Social GeoPref Test – makes use of screen-based infrared (IR) eye-tracking technology and is predictive of ASD and thus useful for early identification, prognosis or subtype specific intervention planning [23].

Earlier this year, in July 2019, Morgan Frost-Karlsson et al published an article about their study using eye-tracking and pupillometry (the measurement of pupil dilation or contraction in response to a study) in a social perception situation of ASD teenagers and young adults. The researchers from the University of Gothenburg, Sweden also used screen-based IR eye-tracking technology in their study which counted with 35 adolescents and young adults with neurodevelopmental disorders, in order to investigate the processing of social and non-social scenes in individuals who meet the DSM-5 criteria for autism and in individuals who do not. They found that the participants with ASD did not show increased pupillary response to scenes containing the image of a human being, contrary to what happened with participants without ASD. Besides that, they also verified that ASD participants were slower to fixate on social elements when presented with social scenes, which correlated with clinical measures of poor social functioning. Their results confirmed that eye-tracking and pupillometric indices are clinically relevant in the field of ASD and they further discussed that analysing changes in visual attention when a social stimulus is present might be an integral part of assessing neurodevelopment [24].

Eye tracking has thus proven to be a valuable tool in cognitive psychology studies in this field.

2.4 UNITY

Unity 3D is the name of a video game engine developed by Unity Technologies in 2005. Since its creation, the engine has been adapted to other uses outside of the gaming industry and is currently used in industries as varied as film, automotive and engineering [25].

In Biomedical Engineering, it has been used to develop various kinds of applications, such as Augmented and Virtual Reality (AR and VR) models and simulations of physiological processes and anatomical structures and simulations of surgical procedures, as well as wheelchair course simulations [26]–[28].

2.5 AUTISM SPECTRUM DISORDER

Autism is a Neurodevelopmental Disorder (ND) that manifests itself very early in a child's life and becomes more and more evident through the years. It is a chronic disorder that requires intervention to minimize its effects [29], [30].

In 2013 a new set of criteria called Diagnostic and Statistical Manual of Mental Disorders (DSM-5) was adopted. Following these criteria, autism is defined as a spectrum (Autism Spectrum Disorder – a definition that was initially proposed by Lorna Wing and Judith Gould in 1979) that encompasses Asperger's disorder, childhood disintegrative disorder, and pervasive developmental disorder not otherwise specified [31].

2.5.1 Comorbidities

ASD has certain comorbidities associated with it. The more frequent ones are, according to DSM-5, Intellectual Developmental Disorder (IDD), Language Disorder (LD) and Attention Deficit and Hyperactivity Disorder. In an intervention perspective it is very important to diagnose these comorbidities so that the planned intervention can be adapted and targeted at the child's specific difficulties [31].

For the purposes of this project, more detail will be provided about LD because the children involved in the study were only diagnosed with this specific comorbidity of ASD.

The most frequent warning sign that appears in children which are later diagnosed with ASD is a delay in linguistic development, which is very often a sign of LD; in some cases, on the other hand, it is simply a delay in the acquisition of language skills associated with a lack of maturity. Therefore, it is extremely important to try to diagnose this comorbidity correctly as soon as the warning signs show themselves; if these signs are ignored and attributed to a simple maturity problem, the lack of a timely intervention in the first few years of the child's life can lead to aggravated symptoms of ASD later in life. Martins and Bandeira de Lima et al. (2013) showed that there is a significant correlation between the linguistic level and the severity of the symptoms of ASD. Billstedt and Gilberg (2005) also showed that individuals diagnosed with ASD who could use complete sentences by the age of 6 had less aggravated symptoms in adulthood [32].

A diagnosis of LD should be considered when the individual presents a significant delay in linguistic development (relative to the reference ages) and a significant and negative dissociation between linguistic and intellectual levels, because not all individuals diagnosed with ASD present signs of LD and, as mentioned before, signs of LD are not enough to give a correct diagnosis. Nevertheless, the diagnosis of LD is important as a comorbidity of ASD, since its presence aggravates the prognosis and intensifies the need for intervention [29].

2.5.2 Evaluation

From a health screening perspective, there are various tools available to evaluate ASD. All of them share the same goal of identifying, as soon as possible, children that show a set of warning signs that warrant further, more formal examination. These screening instruments are questionnaires directed to the parents or caretakers, in which they are asked to rate the child's behaviour with respect to various warning signs such as reduced eye contact, non-responsive behaviour or a refusal of physical contact. Based on this first evaluation through questionnaires, the child may be called in for more formal examination so that the physician can arrive at a diagnosis, using other tools.

Besides evaluation tools for symptoms of ASD, other tools that evaluate the neurodevelopment, the functional profile and the language development should be used when diagnosing the comorbidity of autism.

For these purposes, the most relevant for this study are the Weschler Intelligence Scales for Children (WISC), developed by D. Weschler in 1991. WISC-III is a reference cognitive assessment tool to evaluate intelligence and identify a global level of cognitive ability. The scale that should be used varies with the child's age. Since the participants of this study were between 6 and 10 years of age, the relevant scale is WISC-III, used for subjects for ages between 6 and 17 years. Using this scale, it is possible to obtain a total IQ Score that is in turn comprised of five Primary Index Scores: Verbal Comprehension, Visual Spatial, Fluid Reasoning, Working Memory and Processing Speed.

Verbal Comprehension is an Index Score that measures a child's ability to verbally reason, which can be influenced by their semantic knowledge. The Visual Spatial Index reflects the ability to understand visual details and relationships in order to solve puzzles and construct geometric designs. The Fluid Reasoning Index is related to a child's ability to detect relationships among visual objects. Working memory refers to the ability to register, maintain and manipulate visual and auditory information. Finally, the Processing Speed Index reflects the speed at which a child can accurately make decisions.

This instrument of evaluation consists of sixteen subtests that enable a first distinction between skills or psychological functions, although the standard number of subtests given is only seven. Each one of the Primary Indexes is tied to specific subtests. The subtests used in this project were Coding and Object Assembly, which are tied to the Processing Speed and Visual Spatial Indexes, respectively. [33].

In the clinical applications of these subtests, the scores are given based on total time of execution and for correct positioning of pieces (this last one only applies to the Object Assembly subtest). If the subjects take a short amount of time and make few mistakes, their score is higher. The subject's performance on a subtest is afterwards compared to the normative sample based on their age and then the subtest's scores are converted into scaled scores to be used in the calculation of the total IQ Score. The final score is scaled from 1 to 20 and divided in 6 classes: below average (1-5), low average (6-7), average (8-11), high average (12-13), superior (14-15) and very superior (16-20) [34].

2.6 CONTEXT

In 2016, G. Oliveira et al. from the University of Coimbra performed a study in which they investigated the influence of specific ASD deficits in IQ, Indexes and subtests from WISC-III, using a sample of 445 school-aged children [224 ASD children and 221 children with other neurodevelopment disorders (OND)] matched by Full-Scale IQ and chronological age. The researchers found that ASD children scored lower on the "Comprehension" and "Coding" subtests and higher on the "Similarities". When comparing the subgroup of ASD children without IDD to the subgroup of children with other neurodevelopmental disorders also without IDD, the researchers found once more that ASD children scored lower in the "Coding" subtest. However, in the "Object Assembly" subtest, they found no significant differences [35].

Coding assesses children's ability to hold attention and flexibility in learning and executing an unfamiliar task. This task requires visual scanning and focusing, attentional skills, sequential processing, short-term memory and psychomotor skills. The results of this subtest can be influenced by multiple factors such as anxiety, distractibility, concerns with accuracy and detail, motivation and persistence. Coding can also be sensitive to brain damage but in this experiment this was not a concern since brain damage was an exclusion factor [33].

Object Assembly tests synthesis and integration of components, visual organization, planning ability, simultaneous processing and visual–motor coordination. Performance in this subtest can be influenced by precision of motor activity, ability to respond when faced with uncertainty, previous experience in puzzle solving, flexibility and persistence [33].

In the present study, the intention was to find out if there are significant differences in scoring between ASD children and typical development (TD) children, in both subtests ("Coding" and "Object Assembly"). Coding was selected because it was the task were scores were lower in the study mentioned above. Object Assembly, in this project called Puzzles task, was chosen as the validating task and also to compare ASD and TD populations to verify if there are any differences between these and the populations in that study.

3 MATERIALS AND METHODS

After developing the app, a hypothesis was defined: are there significant differences between TD and ASD children in the execution of the two tasks in the game (Coding and Puzzles), relating to the number of errors, latency time and total time.

3.1 PROJECT IMPLEMENTATION

The first step for this study was the contact was made with Centro Hospitalar Universitário Lisboa Norte to establish a partnership to help develop the app and perform some of the tests. Another important protocol was established with Centro de Neurodesenvolvimento LógicaMentes where most of the tests with children were made. Finally, the study was submitted for Ethics approval at the local faculty.

3.2 PARTICIPANTS SAMPLE

All twenty-four participants in this experiment were selected according to previously defined inclusion and exclusion criteria and divided in two groups: Control – TD, and Experimental – ASD. The inclusion criteria were the following: being between 6 and 10 years of age, being elementary school students and presenting typical psychomotor development (for the control group) and having mild or moderate ASD (according to DSM-5 diagnostic criteria) for the experimental group. Exclusion criteria included severe ASD, sensory deficits (visual and/ or auditory) and IDD.

The participants were selected considering all the criteria above mentioned and availability in the short period of time the tests were conducted. ASD participants were selected from databases at Centro Hospitalar Universitário Lisboa Norte and Centro de Neurodesenvolvimento LógicaMentes.

3.3 MATERIALS

Multiple components were used in the process of developing the app. In the subchapters below, they are described in detail.

3.3.1 Development Platform

Unity 3D was the software selected to develop the app because it is a cross-platform engine that allows the creation of games (in 2D or 3D) and integration of created games with other pieces of software like Tobii 4C eye tracker. The component editing can be done via Unity 3D Inspector (see Figure 3.1) or Visual Studio – one of the C# editors. The Unity 3D version used to develop the app was version 2017.3.1, together with the Tobii Eye Tracking SDK (software development kit) available in the Unity Asset Store [25].

The choice of only using this software to develop the app offers three main advantages: fast and intuitive development language (C#), easy modification or addition of features and facilitating the distribution of the project to the Unity 3D community in open source.

The Unity 3D environment is divided in six sections: “Scene View”, “Game View”, “Hierarchy”, “Project”, “Inspector” and “Console” as shown in Figure 3.1 [36].

- **Scene View** is where the game objects created can be manipulated. This section allows the arrangement and organization of objects but also manipulation of their size. One of the most important objects is the camera because it will transmit all the graphics and audio in the game view.

- **Game View** is rendered from the Camera(s) in the game. It is representative of the final game. One or more Cameras are needed to control what the player sees when they are playing the game.

- **Hierarchy** contains a list of every game object in the current scene. Some of these objects can be empty just to organize the section like “Cards” (Figure 3.1) or 2D, 3D or User Interface (UI). One project can have multiple scenes and the user can switch between them.

- **Project** window allows access and management of all the assets belonging to the project including scenes, scripts, sprites or any other specific files necessary to the project.

- **Inspector** displays detailed information about any selected game object, asset or setting including all attached components and their properties. Some of the information in this window includes the position in the scene, public variables and scripts attached. In the Inspector it is also possible to program some interactions.

- **Console** window displays errors, warnings and other messages generated by Unity. It can be very useful while debugging because it also shows messages programmed by the user.

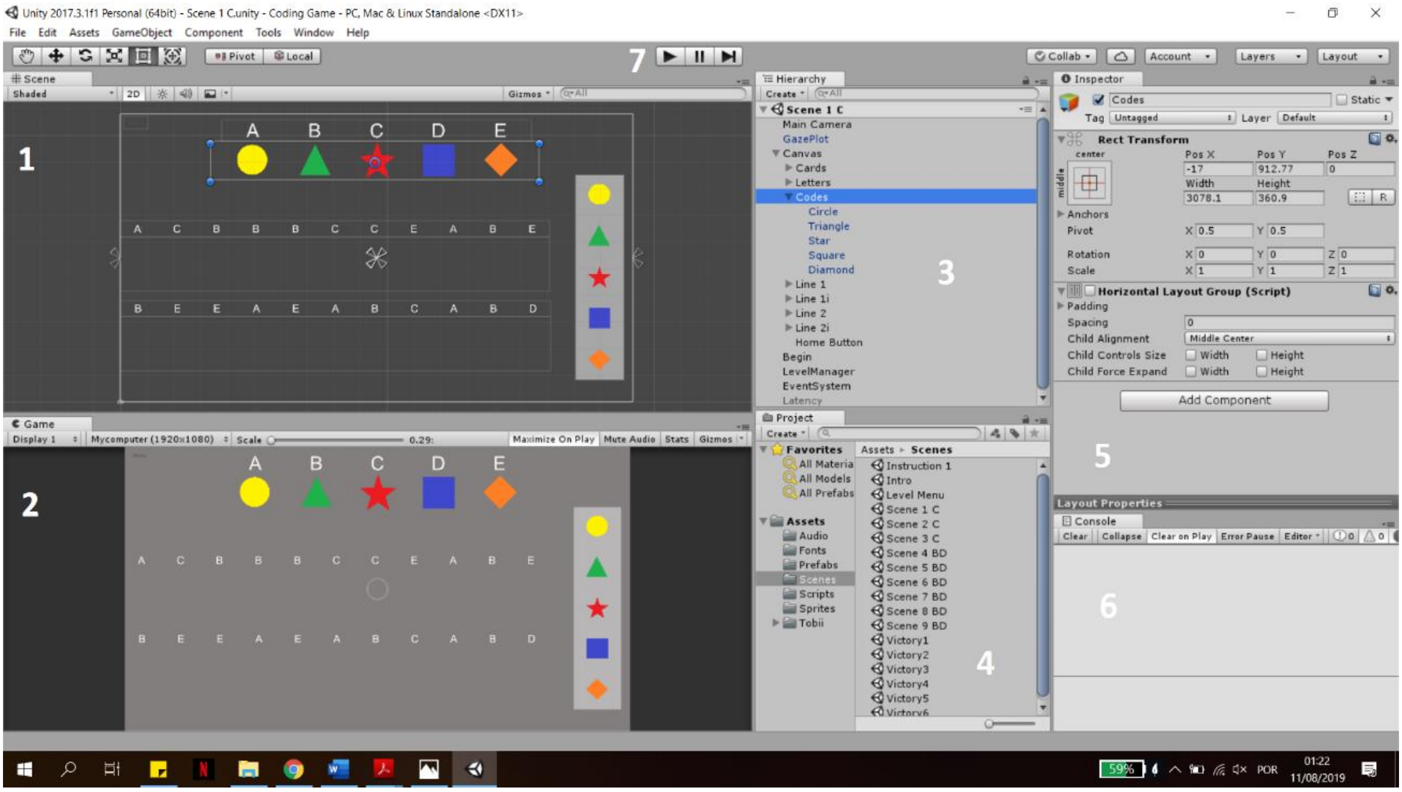


Figure 3.1: Example of a Coding scene in editing mode. Unity 3D is divided in 6 sections: 1 - “Scene View”, 2 - “Game View”, 3 – “Hierarchy”, 4 – “Project”, 5- “Inspector” and 6 – “Console”. To run all the scripts and simulate the game play there are 7 – “Game Buttons”.

3.3.2 Image

One of the most important materials to build this game were the images. A total of 20 images was used throughout the game. There was a preselection of 65 pictures from the website PNGIMG[1] because it has a large library of PNG pictures with a Creative Commons 4.0 BY-NC license which allows usage and adaptation for free [37]. In order to obtain the final selection, the following factors were considered: children interests (assessed by clinical experience with the ASD assessment and intervention, and the pilot test phase with two ASD children), colour, number of elements (in case of landscapes), positioning and size. After considering all these factors, 13 images were chosen to include in the game. The geometric shapes and the **Play** and **Exit** buttons were designed using Paint 3D app and the colours used were selected because they are basic colours within the first six acquired in terms of development of the child.

3.3.3 Sound

The sound elements were used during the game for three purposes. Before each task to explain the task, at the beginning of each scene to give a start instruction, and at the end of each task as a positive stimulus to help keep the children motivated. The game instructions were recorded in .mp3 format and the final applause sound was obtained from SoundBible[2] website also in .mp3 format [38].

---

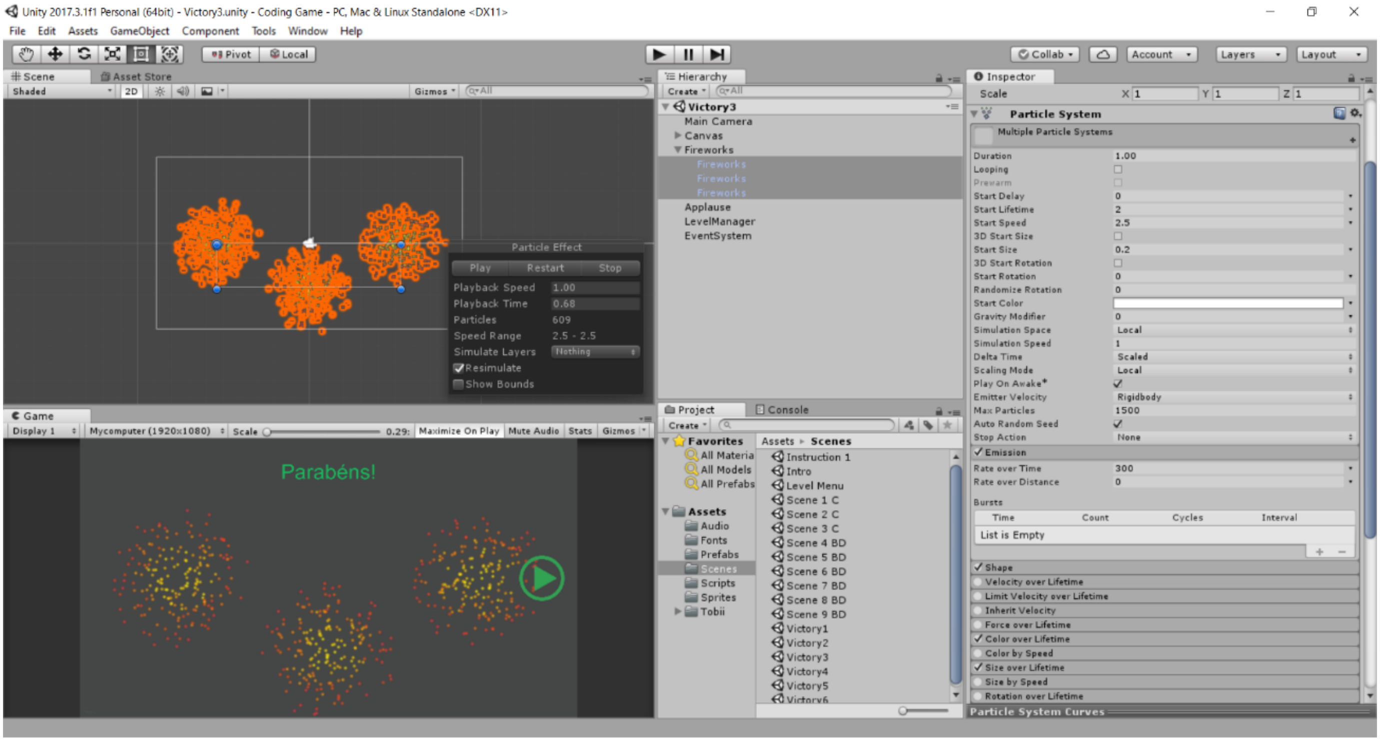
[1] http://pngimg.com/

[2] http://soundbible.com/

3.3.4 Animation

Animations were the last element added to the game. The fireworks animation used in this game was created in Unity 3D as it allows the creation of complex particle systems and adjustment of multiple variables such as duration, simulation speed, shape, colour and size over lifetime (Figure 3.2).

The purpose of including animations throughout the game is to give positive feedback at the end of each task and keep the children motivated to complete the game.


Unity 2017.3.1f1 Personal (64bit) - Victory3.unity - Coding Game - PC, Mac & Linux Standalone <DX11>

Particle Effect
Playback Speed 1.00
Playback Time 0.68
Particles 609
Speed Range 2.5 - 2.5
Simulate Layers Nothing
☑ Resimulate
☐ Show Bounds

Scale X 1 Y 1 Z 1
Particle System
Multiple Particle Systems
Duration 1.00
Looping ☐
Prewarm ☐
Start Delay 0
Start Lifetime 2
Start Speed 2.5
3D Start Size ☐
Start Size 0.2
3D Start Rotation ☐
Start Rotation 0
Randomize Rotation 0
Start Color
Gravity Modifier 0
Simulation Space Local
Simulation Speed 1
Delta Time Scaled
Scaling Mode Local
Play On Awake* ☑
Emitter Velocity Rigidbody
Max Particles 1500
Auto Random Seed ☑
Stop Action None
✓ Emission
Rate over Time 300
Rate over Distance 0
Bursts
Time Count Cycles Interval
List is Empty
✓ Shape
Velocity over Lifetime
Limit Velocity over Lifetime
Inherit Velocity
Force over Lifetime
✓ Color over Lifetime
Color by Speed
✓ Size over Lifetime
Size by Speed
Rotation over Lifetime
Particle System Curves

Figure 3.2: Animation development in Unity 3D.

3.3.5 Eye Tracker

The eye tracker selected to use in this experiment was Tobii 4C, which uses Near Infrared micro projectors (850 nm) to detect eye movement (Figure 3.3), because it offers multiple advantages:

- Portability (17 x 15 x 335 mm^3 and 95 grams) and affordability (150 - 170€);
- Reduced CPU load and power consumption – via Tobii EyeChip™ ASIC;
- Simple integration – easy to setup and start using in any PC monitor or laptop. It's connected through USB 2.0 and has support in Windows 10;
- Simultaneous eye and head tracking;
- Easy integration in Unity 3D projects;
- Free SDK with a large list of functions.

As mentioned before, there are different kinds of eye trackers. Screen-based and wearable eye trackers use the same technology. However, wearable devices can be disturbing for ASD children, therefore a screen-based device was selected.

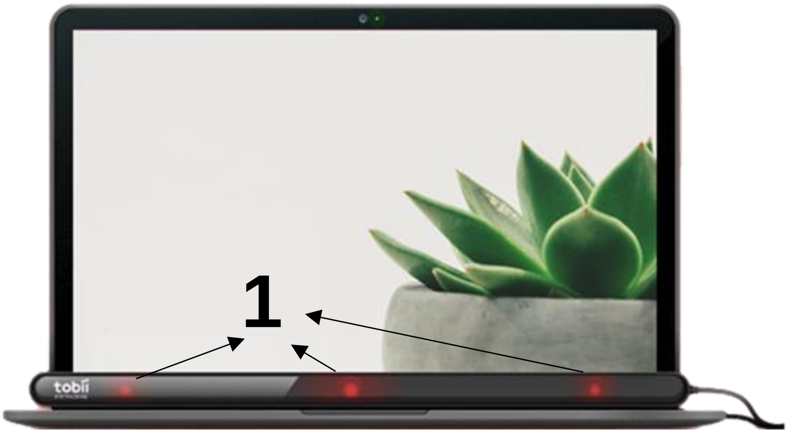


Figure 3.3: Tobii 4C eye tracker setup in laptop. The three red points (1) are the NIR micro projectors.

The Tobii 4C device was used to collect eye tracking data (XY coordinates) at a frequency of 60 Hz and an operating distance between 50 and 95 cm. Figure 3.3 demonstrates the eye tracker setup in a laptop as used in this study.

At the start of each test, a calibration was performed for every participant using the calibration procedure included in the base software. This procedure is composed by 7 points. The participant must look at each dot until in explodes and the next set of dots appears. Figure 3.4 exemplifies a calibration screen with three dots.

The biggest limitation in the use of this equipment was the difficult and very expensive access to the Pro SDK licenses. Tobii Pro SDK gives access to Analytical Use in applications, extended data (gaze data for both eyes individually, pupil data if applicable and 3D eye coordinates – gaze origin) and compatibility with Tobii Pro Lab software [39]. This considerably limited the type of data analysis performed in this study.

Figure 3.4: Example of calibration screen. The first screen presents only 1 dot. In this there are 3 dots to pop. After all the dots are popped another screen with 3 final dots appears.

3.4 APP DEVELOPMENT

The app developed for this project is a game with two tasks integrated with the eye tracker software which collects data in the background as the participant plays the game. As mentioned before, the tasks developed were Coding task and Puzzles task. These were designed based on two performance subtests from the WISC-III.

The tasks included in this project are based on the performance subtests: Coding and Object Assembly. For each task there are three levels with increasing difficulty.

The app was built with Windows as the target platform (x86_64 architecture) and programmed to use the mouse as the input device. The game runs in the following order:

1. Initial screen;
2. Instruction scene;
3. Coding task – level 1;
4. Victory scene;
5. Coding task – level 2;
6. Victory scene;
7. Coding task – level 3;
8. Victory scene separating tasks (fireworks and applause);
9. Puzzles task – Image 1;
10. Puzzles task – Level 1;
11. Victory scene;
12. Puzzles task – Image 2;
13. Puzzles task – Level 2;
14. Victory scene;
15. Puzzles task – Image 3;
16. Puzzles task – Level 3;
17. Final Victory scene (fireworks and applause).

In every level there is always the possibility to return to the level menu and exit the game. This option was especially thought for ASD children that can easily get tired of a task or need to stop the game.

3.4.1 Coding Task

As mentioned before, this task was designed based on the Coding subtest of the WISC-III. The goal of this game is to observe the key, composed by 5 letters (A to E) paired with different figures, and then match those figures to the letters in random sequences as Figure 3.5 demonstrates. The sequences were randomly generated via Microsoft® Office Excel (Annex A – Figure 3.13) and have eleven letters each in every level.

In order to complete the level, the participant must click and drag the matching figure to the space under the letter, one by one, until every letter has the matching figure below. The participants were instructed to complete one entire line before advancing to the next. If the combination is correct, the image is locked in place under the letter. Otherwise, the image returns to the inventory (Figure 3.5 - 2) and the player must try again. There is no limit of attempts per letter and no limit of time to complete each level.

This test is designed to evaluate speed of processing, but it also measures short-term memory, concentration, non-verbal learning and visual-motor dexterity.

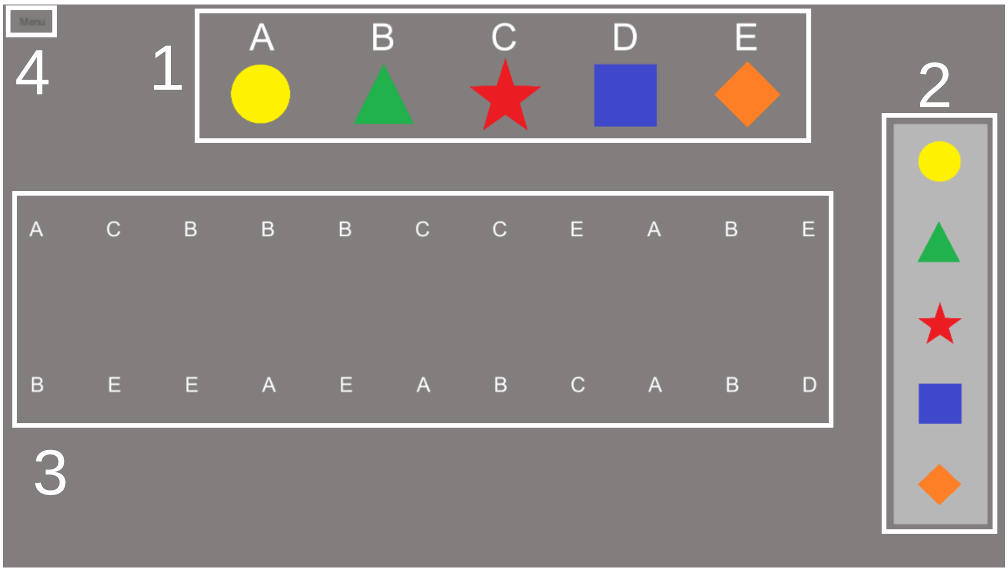
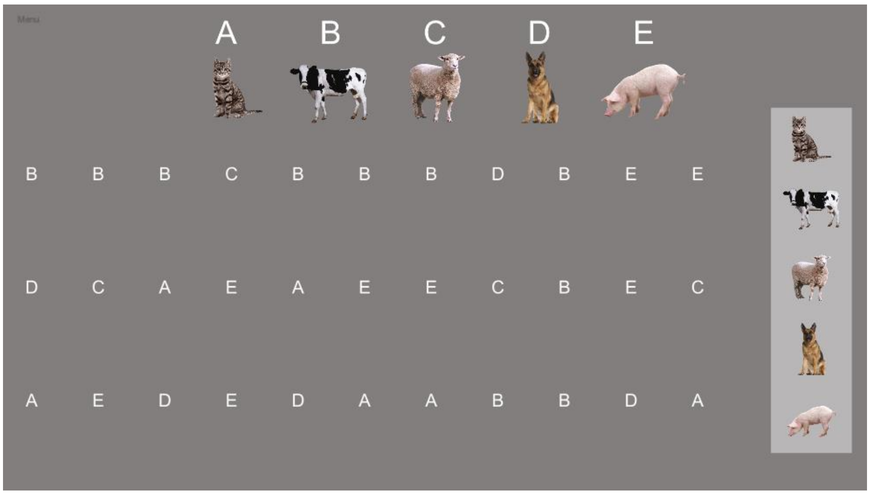


Figure 3.5: First level of Coding task. The scene is composed by three elements: 1 – key: 5 letters paired with geometric figures, 2 – inventory: figures to drag to the matching letters in random sequences, 3 - random sequences of letters, 4 – **Menu** button.

There are three levels for the Coding task. For each level there is a different key determined by the images selected: level one – geometric figures (Figure 3.5: circle, triangle, star, square and diamond), level two – farm animals (cat, cow, sheep, dog and pig), level three – land vehicles (car, bicycle, bus, motorcycle and truck). Additionally, in each level the number of lines with random sequences is increased: level one – two lines, level two – three lines, level three – four lines. Considering these changes in every level it is possible to gradually increase the difficulty of the task while keeping children motivated by changing the images in each scene.

Also, between each level there is a Victory scene to indicate the end of the level and give positive feedback.



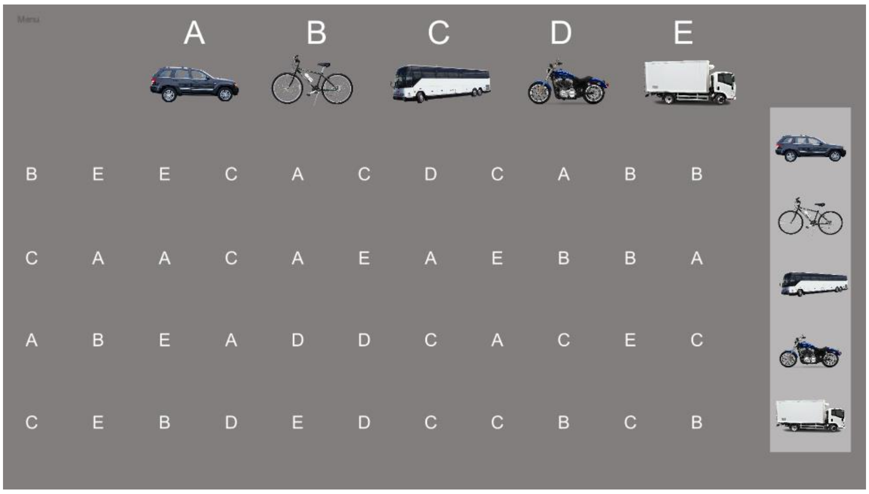
A B C D E

B B B C B B B D B E E

D C A E A E E C B E C

A E D E D A A B B D A

Figure 3.6: Coding task level two - Farm animals.



A B C D E

B E E C A C D C A B B

C A A C A E A E B B A

A B E A D D C A C E C

C E B D E D C C B C B

Figure 3.7: Coding task level three - Land transports.

3.4.2 Puzzles Task

The Puzzles task was also based on a subtest from the WISC-III: Object Assembly. The goal of this game is to observe an image for thirty seconds and afterwards build the puzzle. Figures 3.6 and 3.7 demonstrate two examples of scenes that compose this task. In order to complete the level, the participant must drag each piece of the puzzle to the correct slot until the image previously presented is rebuilt. To see all the pieces available in the inventory the player uses the mouse scroll wheel. In order to define the order of the pieces in the inventory, once again random sequences were generated using Microsoft® Office Excel (Annex A – Tables 3.3, 3.4 and 3.5). If the piece is dropped into the correct slot, it is locked in place. However, if the piece is placed incorrectly it will return to the inventory. The participant can choose to pick up the same piece again or try to place another one. There is no limit of attempts to place a piece and no time limit to complete the level.

This test is built to evaluate the ability to analyse and synthesize an abstract design and reproduce it. At the same time, the task measures memory, perceptual organization and reasoning with part/whole relationships, spatial visualization, simultaneous processing, visual-motor coordination, visual-motor dexterity, and nonverbal concept formation.

Figure 3.8: Puzzles task level one picture. The picture is presented for 30 seconds. When the time is up, the puzzle automatically appears.

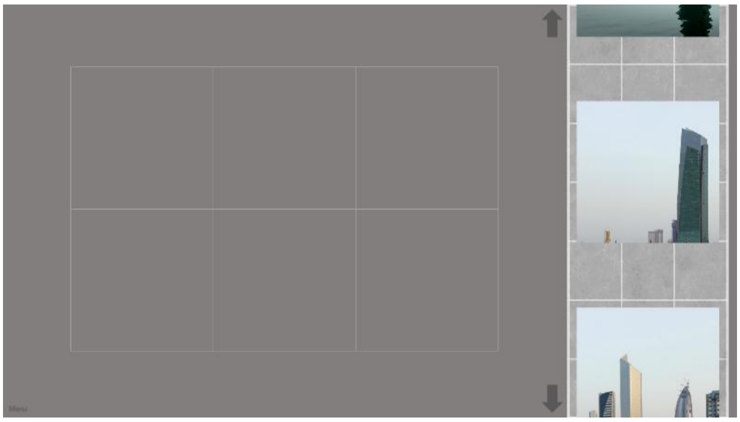
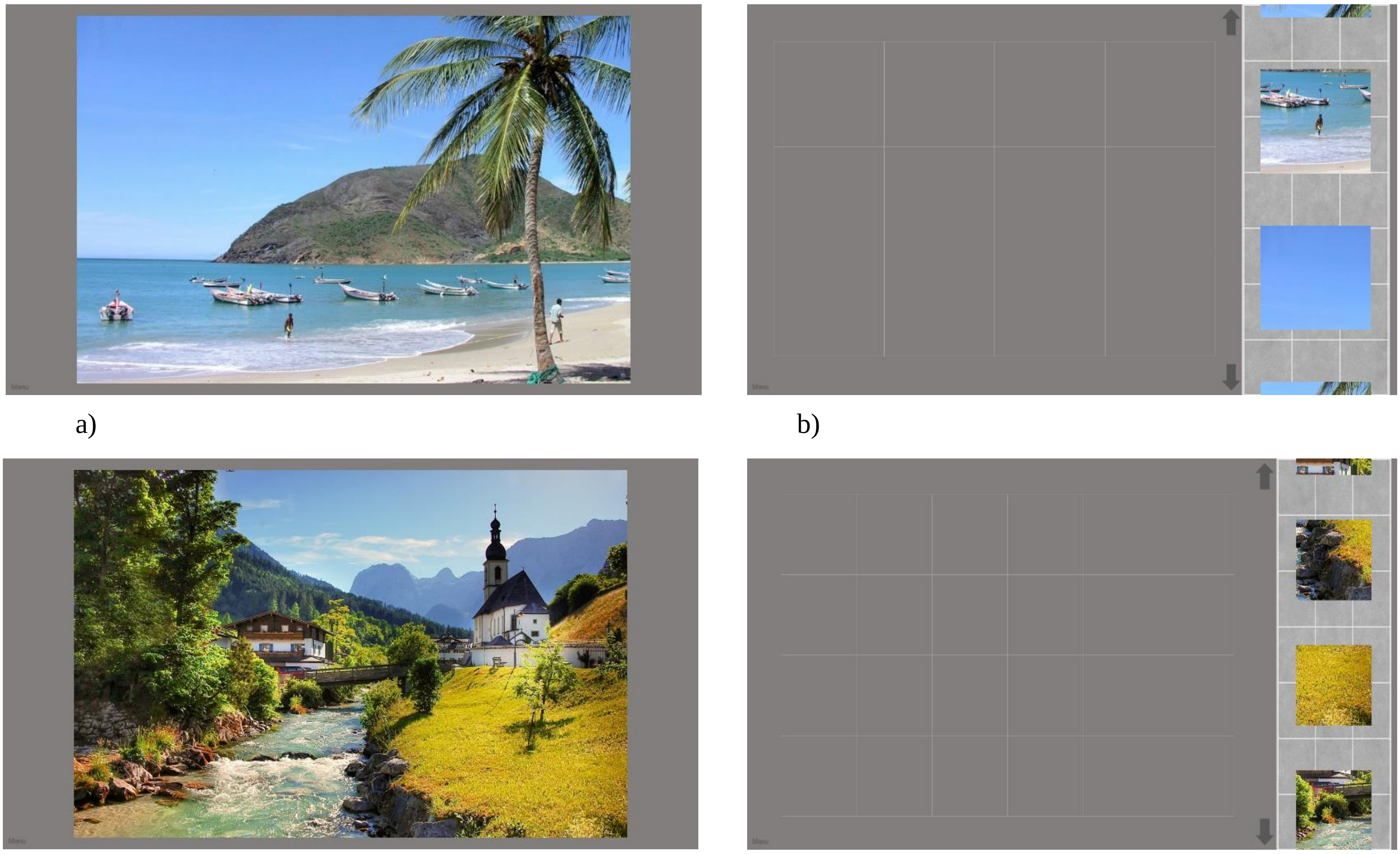


Figure 3.9: Puzzles task level one with 6 pieces. The white grid helps the child know where to place the pieces and complete the task.

As in the Coding task, there are also three levels for the Puzzles task. Each level has a different picture of landscapes: level one – city (Figures 3.8 and 3.9), level two – beach (Figure 3.10 a)), level three – mountains (Figure 3.10 c)). These three images were picked considering two important factors: the number of elements in the picture and the colour distribution. Additionally, in each level the number of puzzle pieces increases: level one – six pieces (Figure 3.9), level two – twelve pieces (Figure 3.10 b)), level three – twenty-four pieces (Figure 3.10 d)). Considering these changes, the difficulty of the task increases gradually while keeping children motivated. Also, between each level there is a Victory scene to indicate the end of the level and give positive feedback.

The three landscapes pictures were selected to assess the children’s ability to reconstruct a complete image. Introducing face pictures would create an added difficulty for the ASD participants.



a) b) c) d)

Figure 3.10: Puzzles task levels: a) and b) - level two; c) and d) - level three.

3.4.3 Eye Tracker Integration

The last part of the app development was integrating the eye tracker software with the game. To do so, the only software used was the Tobii Eye Tracking SDK to program the eye tracker to run in the background, while the participant plays the game, and collect eye movement data.

Because of the restrictions previously mentioned, the only data collected were the X and Y coordinates of the gaze point. In order to obtain these coordinates, the script written compares the most recently saved gaze point time stamp with the previous gaze point time stamp, and if it is higher, the coordinates are saved to a .csv file. As this is always true, each frame a new pair of XY coordinates is saved.

3.5 TESTING PHASE

The testing phase was a very important part of the project as a proof of concept. The main goal was to understand the practical potential of the app as an evaluation tool for the executive functions of children with ASD compared with TD children.

Besides the results obtained from the game execution, the behaviour during the test was registered and a short survey was done at the end using Google Forms. This questionnaire was a helpful tool to learn more about the children's acceptance of the game, which task they preferred and which task was the hardest. The results of those surveys are presented in Annex B).

All the tests were conducted in the institutions with established protocols: Centro de Neurodesenvolvimento Infantil LógicaMentes and Centro Hospital Universitário Lisboa Norte.

3.5.1 Participants

Since all the participants in the experiment were minors, their legal guardians were provided an information leaflet and required to read and sign a written informed consent form before the test (Annex C).

A total of sixteen TD children participated in the experiment. All subjects were either friends' or work colleagues' children who agreed to be part of the study voluntarily. The subjects were 11 females and 5 males with an average age of 7.6 ± 1.2 years (see Table 3.1). In Table 3.1, two participants are marked with an (*) because they were the subjects to do the pilot test.
The experimental group was composed by eight ASD children carefully selected according to the inclusion criteria. All the participants' legal guardians agreed for their children to be part of the study voluntarily. The subjects were 2 females and 6 males with an average age of 7.9 ± 1.5 years (see Table 3.2).

Table 3.1: Overview of typical development participants' characteristics.

Typical Development Subjects	**Gender**	**Age (years)**	**Handedness**
1	Male	7	Right
2*	Female	7	Right
3	Female	6	Right
4	Male	8	Right
5	Female	6	Right
6	Female	7	Right
7	Female	9	Right
8	Female	8	Right
9	Female	8	Right
10	Male	7	Right
11	Female	7	Right
12	Female	6	Left
13*	Female	8	Right
14	Female	8	Right
15	Male	10	Right
16	Male	9	Left
Mean ± std		7.6 ± 1.2	

std standard deviation; * Pilot test participants.

Table 3.2: Overview of Autism Spectrum Disorders participants' characteristics.

Autism Spectrum Disorders Subjects	Gender	Age (years)	Handedness	Comorbidities
1	Male	9	Right	LD
2	Male	6	Right	LD
3	Male	8	Right	LD (non-verbal)
4	Male	6	Right	LD (non-verbal)
5	Male	7	Right	–
6	Male	10	Right	–
7	Female	8	Right	LD
8	Female	9	Right	–
Mean ± std		7.9 ± 1.5		

std standard deviation

3.5.2 Procedure

The experimental procedure established for TD and ASD children was the same. It is also important to mention that none of the participants had any previous experience with the app.

All the participants in the study performed a single session of testing. The session's duration was between thirty minutes to one hour depending on factors such as the child's overall behaviour, acceptance of the game (especially with ASD children), distraction factors or difficulties in execution. The parents/legal tutors had the choice to stay in the room during the test or wait outside.

Before every session started, the computer and eye tracker were setup and connected to make sure everything was functioning properly. The experimental protocol was defined with the following steps:

1. Send the information leaflet before the test;
2. Reading and signing of the informed consent form after clarifying any existing questions;
3. Save child's data: age, school year and handedness;
4. Brief procedure and experiment description and explaining the eye tracker function;
5. Instruct the participant to sit comfortably but straight and move as little as possible during the test – the chair was facing the wall during the test to reduce distraction factors;
6. Confirm eye tracker connection, explain calibration procedure and start calibration;
7. Launch the game. If necessary, reinforce orally the instructions given in the game;
8. After completing the game, ask final questions and fill in the survey;

9. Give certificate of participation (Annex D) to the child and thank the parents/legal tutors for their availability.

3.5.3 Pilot Test

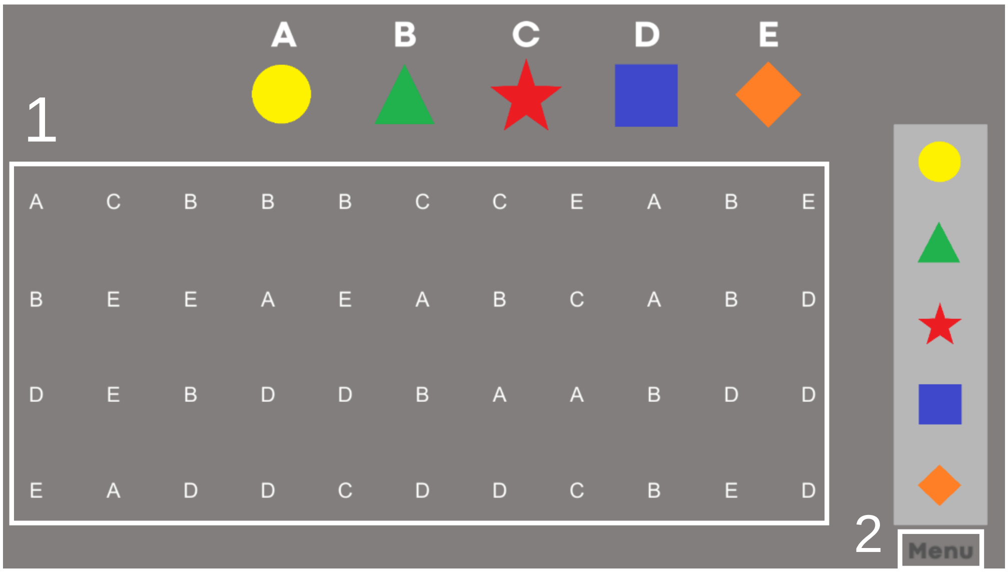
The first two trials were used as a pilot test to detect any errors previously unnoticed and make some final adjustments, if necessary. The two children who completed this pilot test belonged to the control group.

After observing the behaviour of the participants during the session and analysing the results obtained from these tests, some changes were made to the first version of the app. In this version, all the Coding task scenes had four rows of random sequences of letters, as demonstrated by Figure 3.11 (1). The letters (A, B, C, D, E) were selected because they are the first five letters of the alphabet and children within these range of ages are already aware of them. Also, ASD children have a specific interest in this kind of stimulus, according to clinical experience.

In order to have more distinction in difficulty between each level and to reduce the fatigue exhibited by children during the pilot test, the number of lines was reduced to two in the first level and three in the second level (final version described in Chapter 3.3.1).

Another small change made was the **Menu** button positioning and size (Figure 3.11 - 2). After the pilot test, the positioning was considered wrong because it could lead to some accidental clicks. Also, the button's size was considered a distraction factor. In the final version it was repositioned on the top left corner as illustrated in Figure 3.5.

To make the game less tiresome, the hypothesis of alternating a Coding scene with a Puzzle scene was considered but the final version kept the same order as the original to try to verify if there was any progression in task execution in three consecutive levels.



A	C	B	B	B	C	C	E	A	B	E
B	E	E	A	E	A	B	C	A	B	D
D	E	B	D	D	B	A	A	B	D	D
E	A	D	D	C	D	D	C	B	E	D

Figure 3.11: First level of the Coding task in the first version of the app, used in the pilot test: 1 - 4 lines of random sequences; 2 - Menu button.

3.6 MEASURES

For each level in the game, three measures were recorded: number of errors, latency time and total time of execution. This was called the Score data. The eye tracker collected data simultaneously in background, called the Gaze data. For every participant, the data was saved in files named by scene name and separated in folders according to the data classification.

According to the EU General Data Protection Regulation (GDPR), all subjects' data was anonymized and saved in password protected files [40].

3.6.1 Score Data

During the execution of the game, these three measures were registered per level for each subject. In the sub-chapters below, they are described in detail. The data collected was not visible or accessible to participants at any time during or after the game.

All the statistic results obtained from these measures were analysed using R which is a free software environment for statistical computing and graphics. This software is available for free under the terms of the Free Software Foundation's GNU General Public License (GPL) [41]. As R is an integrated set of software facilities for data manipulation, calculation and graphical display, it offers several advantages:

- Practical data handling and storage facility;
- Large list of operators for calculations on arrays;
- Intermediate tools collection for data analysis;
- Graphical facilities for data analysis and display;
- Simple and effective programming language.

The results will be present in detail in the next chapter.

3.6.1.1 Number of Errors

This measure refers to the total number of errors per level.

In the Coding task, an error is counted every time the player drops the wrong image under a letter and the image returns to the inventory. As there is no limit of attempts per letter, there is also no limit for errors. If the image is held and released in any other area in the screen besides under the sequences, it is not considered an error.

For the Puzzles task, an error is counted whenever the player tries to place a piece in the wrong area. If the player grabs and releases a piece in the inventory or any area outside of the grid it is not considered an error.

This method of error counting can have a small margin of error in cases where the picture is accidentally released in the wrong place or in cases of participants with very little experience with computer mouse that don't have a lot of control when positioning the images.

3.6.1.2 Latency Time

The latency time was defined as the time between the start of the level and the first click and drag of a picture (Coding task) or puzzle piece (Puzzles task). In case the participant clicks any other part of the screen the latency time keeps increasing.

This measure is important to determine how long the child requires to focus and start the task.

3.6.1.3 Total Time of Execution

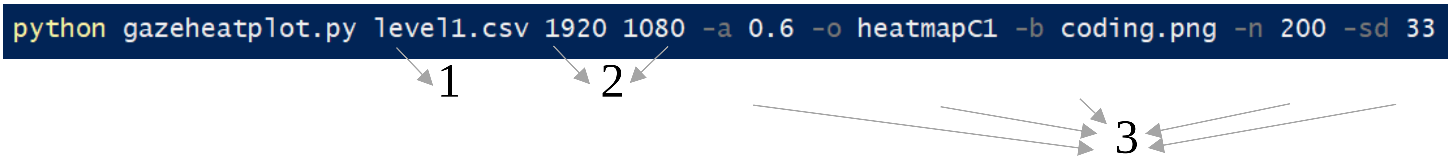
This measure was defined as the time passed since the start of the level until the moment the player places the last image (Coding task) or puzzle piece (Puzzles task).

The time of execution indicates the difficulty in understanding and completing the task, although it can be influenced by the participant's dexterity, concentration and ability to deal with frustration.

3.6.2 Gaze data

As previously mentioned, the gaze data collected were the XY coordinates. These coordinates were later used to generate heatmaps, using a Python command line-based tool by Tobias Roeddiger which is licensed under the GNU GPL v3.0 [42].

The version of Python used with this tool was Python 2.7. In order to obtain the final map for each dataset it is only necessary to run the script from the command line demonstrated in Figure 3.12. There are three required arguments: input path (1), display width and height (2). Additionally, it is possible to adjust the optional arguments (3): alpha of the gaze overlay (-a), output file name (-o), path to background image (-b), gaussian matrix width and height (-n) and gaussian distribution standard deviation (-sd).



```
python gazeheatplot.py level1.csv 1920 1080 -a 0.6 -o heatmapC1 -b coding.png -n 200 -sd 33
```

Figure 3.12: Command line that runs the Python script and generates the heatmaps. Required arguments: 1 – input path; 2 – display width and height. 3- Optional arguments.

This type of maps allows an initial qualitative analysis of the eye tracking data and comparison of areas of interest between the two groups. The resulting heatmaps, analysis and discussion are detailed in the next chapters.

3.7 DATA ANALYSIS

The final step in this project was the statistical data analysis. As mentioned before, R was the software chosen to perform the tests described in this subchapter.

Another analysis was to calculate the average time per piece/ letter over levels for both groups. To do so, the first step was calculating this time for each subject and then calculate the mean.

3.7.1 Boxplot

There are some studies in which it necessary to have information on the variability or dispersion of the data. A boxplot is a graph that gives a good indication of how the values in the data are spread out.

Boxplots are a convenient tool to display data distribution through quartiles and have the advantage of using less space, which is useful when comparing distributions between many groups or datasets. These plots can be drawn horizontally or vertically.

The lines extending parallel from the boxes are known as the “whiskers”, which are used to indicate variability outside the upper and lower quartiles. When the data contains outliers, these are sometimes plotted as individual dots that are in-line with whiskers [43].

This type of plots was used in the analysis of the results to compare the age distribution of the two groups studied.

3.7.2 Fisher’s Exact Test

Fisher's exact test is a statistical significance test used in the analysis of contingency tables for two nominal variables. This test is normally employed when sample sizes are small, although it is valid for all sample sizes [44].

It is included in the class of exact tests because the significance of the deviation from a null hypothesis can be calculated exactly, rather than relying on an approximation that becomes exact in the limit as the sample size grows to infinity, as with many statistical tests.

To perform this test there are two assumptions:

- The observations constitute an independent sample;
- Each observation can be classified in only one category. The categories are mutually exclusive.

In this experiment, this test was used to verify if the gender differences between groups were significant.

3.7.3 Permutation test

Permutation tests are a subcategory of non – parametric statistics. This test is designed to determine whether the observed difference between the sample means is large enough to reject, at some significance level, the null hypothesis [45].

These tests have multiple advantages. They exist for any test statistic, regardless of whether or not its distribution is known, allowing to choose the statistic which best discriminates between hypothesis and alternative and which minimizes losses. Permutation tests can also be used for analysing unbalanced distributions and combining dependent tests on mixtures of categorical, ordinal, and metric data. Finally, these may be ideal for analysing quantitative data that does not satisfy statistical assumptions underlying traditional parametric tests [46].

In this study, the permutation test was used to compare the means of the number of errors from the two groups.

3.7.4 Mann-Whitney test

The Mann-Whitney U test is considered the nonparametric equivalent of the t-test for two independent samples. This test is used to compare two independent populations relatively to their location. More precisely, it compares the medians of the two populations [47].

The only requirements for this test are that the data should be, at least, in ordinal scale and the variables are continuous.

This test was used to compare the ages, latency time and total time of execution between the two groups.

3.7.5 Spearman’s Rank Correlation Coefficient

Spearman’s rank correlation coefficient (ρ) is a nonparametric measure of rank correlation and measures the strength and direction of association between two ranked variables. It is appropriate for continuous and discrete ordinal variables.

Spearman's correlation assesses monotonic relationships (linear or not). The Spearman correlation coefficient can assume values from +1 to -1. A ρ of +1 indicates a perfect association of ranks, a ρ of zero indicates no association between ranks and a ρ of -1 indicates a perfect negative association of ranks.

Spearman’s ρ was calculated to compare the association between the three measures of interest in this experiment: number of errors, latency time and total time of execution.

4 RESULTS

In this chapter all the results from the Testing Phase described in Chapter 3 are presented in three subchapters: behaviour, score and gaze.

Before analysing the results from the test, a group analysis was performed to better understand the age distribution of the participants. To better visualize these differences between groups, boxplots were used as demonstrated in Figure 4.1.

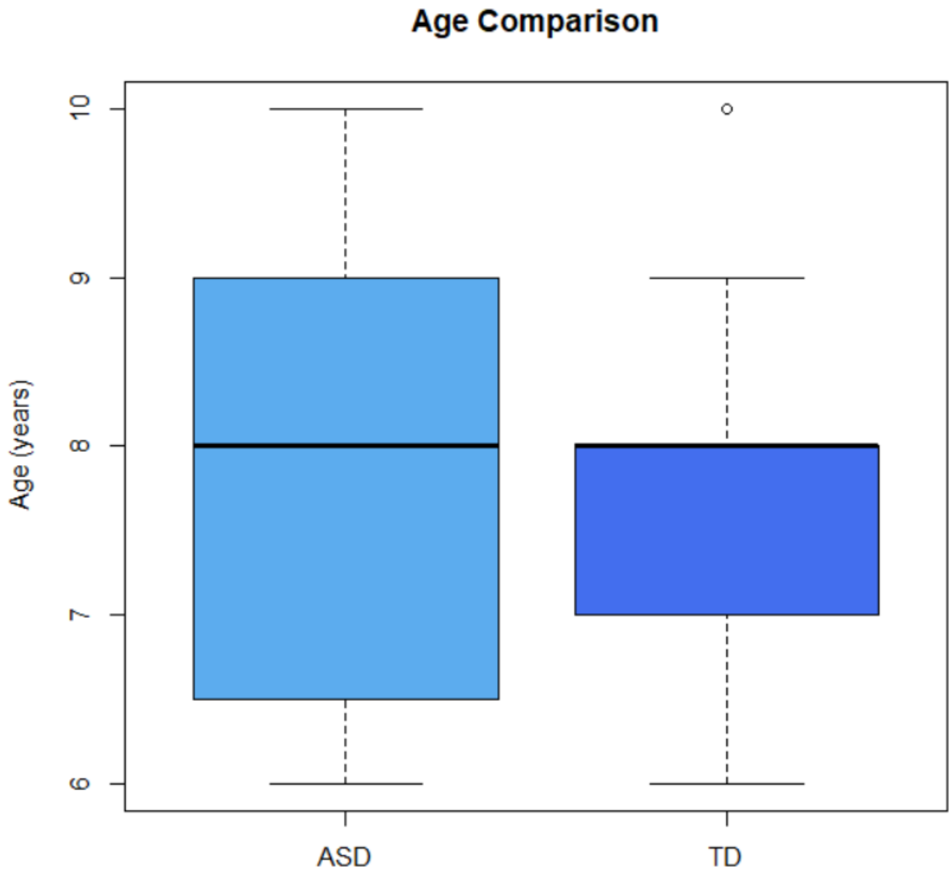


Figure 4.1: Age comparison plots for TD and ASD participants.

The median age for both groups was 8 years. However, it is possible to observe the distributions are very different. The control group distribution is more uneven than the experimental group. This means the TD group had more children with ages between 7 and 8 years. In the ASD group most children had between 6 and 9 years. Nonetheless, there were no significant differences between groups concerning age (Mann-Whitney U test: *p* – value = 0.349).

In regard to gender, there were also no significant differences between the two groups (Fisher's exact test: p – value = 0.183). However, looking at the samples it is clear that number of feminine subjects in the control group is much larger than in the experimental group.

4.1 SCORE RESULTS

4.1.1 Typical Development Group

All fourteen participants in this group completed all the levels for both tasks in the game (100% task completion). In Table 3.1 and 3.2 the results obtained are summarized.

In the Coding task, the average number of errors per scene was 0.86, 0.50 and 1, and the maximum per scene was 6, 4 and 5 as Table 3.1 shows. The average time of execution increased for each scene, as expected, since the number of sequences also increased in each level. The average latency time was higher in level 2 (7.7 s) and was lowest in the first scene (4.4 s).

Table 4.1: TD children score results for the Coding task.

	Scene 1			**Scene 2**			**Scene 3**		
	E	LT (s)	TT (s)	E	LT (s)	TT (s)	E	LT (s)	TT (s)
Mean	0.86	4.4	165.8	0.50	7.7	249.9	1.00	6.8	323.7
Median	0.00	4.0	123.8	0.00	8.3	206.3	0.00	6.5	327.6
Std	1.83	2.6	87.2	1.09	3.4	130.6	1.58	2.3	116.9
Range	0 – 6	1.6 – 11.6	74.2 – 328.1	0 – 4	1.3 – 12.3	102.1 – 636.0	0 – 5	3.1 – 9.5	131.1 – 559.2

E – number of errors; LT – latency time; TT – total time of execution

In the Puzzles task, the average number of errors in each scene was 2, 8.14 and 20.14, and the maximum was 5, 17 and 40 as Table 4.2 shows. The average time of execution increased for each scene, as expected, since the number of pieces also increased in each level. The lowest average latency time was in the second scene (4.9 s).

Table 4.2: TD children scores for the Puzzles task.

	Scene 1			**Scene 2**			**Scene 3**		
	E	LT (s)	TT (s)	E	LT (s)	TT (s)	E	LT (s)	TT (s)
Mean	2.00	6.8	85.4	8.14	4.9	183.7	20.14	6.7	435.8
Median	2.00	6.6	85.8	8.50	4.5	183.9	17.00	6.4	420.5
Std	1.36	2.9	26.3	5.14	2.6	75.7	10.23	1.9	159.4
Range	0 – 5	2.7 – 12.3	42.8 – 128.8	0 – 17	1.9 – 12.2	56.3 – 350.3	8 – 40	3.5 – 9.6	208.1 – 750.5

E – number of errors; LT – latency time; TT – total time of execution

4.1.2 Autism Spectrum Disorders Group

In this group the completion of the game was harder for several reasons mentioned before. The Puzzles task was completed by seven of the eight participants in this group (87.5% task completion). However, the Coding scenes were only totally completed by five participants (62.5% task completion). Levels one and two were completed by six participants but only five children completed level three.

In the Coding task, the average number of errors per scene was 2.14, 2 and 3.60. The maximum number of errors per scene was 8, 8 and 9 as Table 4.3 shows. The average time of execution also increased for each scene, as expected. The latency time was the only measure that did not follow any pattern and was lowest in scene two (7.4 s).

Table 4.3: ASD children score results for the Coding task.

	Scene 1			**Scene 2**			**Scene 3**		
	E	LT (s)	TT (s)	E	LT (s)	TT (s)	E	LT (s)	TT (s)
Mean	2.14	11.9	257.4	2.00	7.4	325.6	3.60	11.9	457.5
Median	1.00	5.2	232.4	1.00	7.6	325.1	3.00	9.7	461.5
Std	2.97	10.9	97.5	2.77	2.8	85.0	3.29	9.5	118.8
Range	0 – 8	2.8 – 30.7	134.4 – 401.1	0 – 8	4.0 – 11.2	222.5 – 469.2	0 – 9	2.3 – 27.5	340.5 – 645.5

E – number of errors; LT – latency time; TT – total time of execution

In the Puzzles task, the average number of errors per level was 2.25, 8.14 and 29. The maximum per scene was 7, 21 and 41 as Table 4.4 shows. The average time of execution also increased for each scene, as expected. The latency time was the only measure that did not follow any pattern and was lowest in scene two (7.0 s).

Table 4.4: ASD children scores for the Puzzles task.

	Scene 1			**Scene 2**			**Scene 3**		
	E	LT (s)	TT (s)	E	LT (s)	TT (s)	E	LT (s)	TT (s)
Mean	2.25	19.7	92.0	8.14	7.0	181.7	28.86	24.7	514.8
Median	1.50	8.3	84.8	8.00	6.9	166.5	33.00	7.2	535.6
Std	2.77	28.7	51.9	7.06	3.5	89.3	12.56	48.8	152.2
Range	0 – 7	4.4 – 89.1	29.3 – 197.9	0 – 21	3.0 – 12.0	69.2 – 339.5	7 – 41	3.4 – 135.3	299.1 – 738.0

E – number of errors; LT – latency time; TT – total time of execution

4.1.3 Group Comparison

The first comparison between groups is the completion percentage of each game task. Looking at Table 4.5 it is clear that the ASD participants struggled to complete the game, especially the Coding task, which was the most challenging for this group.

Table 4.5: Task completion percentage for both groups.

	TD	**ASD**
Coding Task	100%	62.5%
Puzzles Task	100%	87.5%

In order to compare the differences between TD and ASD group by task level, two tests were performed: permutation tests and Mann-Whitney tests. The number of errors is a discrete variable therefore permutation tests were used in this case. The latency and total time are continuous variables, so Mann-Whitney tests were used for these measures.

In Table 4.5 the results for the permutation tests are presented. Only one of the comparisons done showed significant statistical differences, in the Coding task level 3 (p – value = 0.039) for a confidence interval of 95%.

Table 4.6: Significance levels of Permutation tests for number of errors.

Task	Level	Number of Errors (p – values)
Coding	Level 1	0.224
	Level 2	0.088
	Level 3	0.039*
Puzzles	Level 1	0.769
	Level 2	0.988
	Level 3	0.103

Permutation tests: *p < 0.05. All comparisons signalled with * are significant.

In Table 4.7 the results of the Mann-Whitney tests for two independent samples are presented. In latency time there were no significant statistical differences (U test $p > 0.05$) for a confidence level of 95%. In total time of execution there were significant statistical differences for all the levels in Coding task ($p = 0.015$, $p = 0.016$ and $p = 0.009$). In the Puzzles task total time of execution showed no significant statistical differences (U test p > 0.05).

Table 4.7: Significance levels of Mann-Whitney U tests for latency time and total time of execution.

Task	Level	Latency Time (p – values)	Total Time (p – values)
Coding	Level 1	0.064	0.015*
	Level 2	0.655	0.016*
	Level 3	0.104	0.009*
Puzzles	Level 1	0.165	0.513
	Level 2	0.112	0.656
	Level 3	0.514	0.161

U tests: *$p < 0.05$. All comparisons signalled with * are significant and related to superior time results in the ASD group.

Another tool used for this type of analysis were line charts. These graphics were built to better visualize and compare progression over levels for both tasks. The charts were produced considering the mean and median of each measure of interest – number of errors, latency time and total time of execution.

The means are represented by the solid lines and the medians by the dotted lines. Because means are much more sensitive to outliers than medians and the sample size is quite small, it is better to focus on the median differences between groups.

The number of errors progression per group (TD – blue lines, ASD – yellow lines) for the Coding task is represented in Figure 4.2. Although the statistical analysis did not show any significant differences between groups, by observing Figure 4.2 the differences are noticeable. Both groups have similar progression considering the mean but by analysing the median lines it is possible to verify the ASD group had considerable progression in level 3, while the TD group did not change across levels.

The number of errors progression per group for the Puzzles task is represented in Figure 4.3. In this case the mean and median are very close and the only noticeable differences between groups are in level 3. The TD group progression is almost linear while the ASD values increase drastically in level 3.

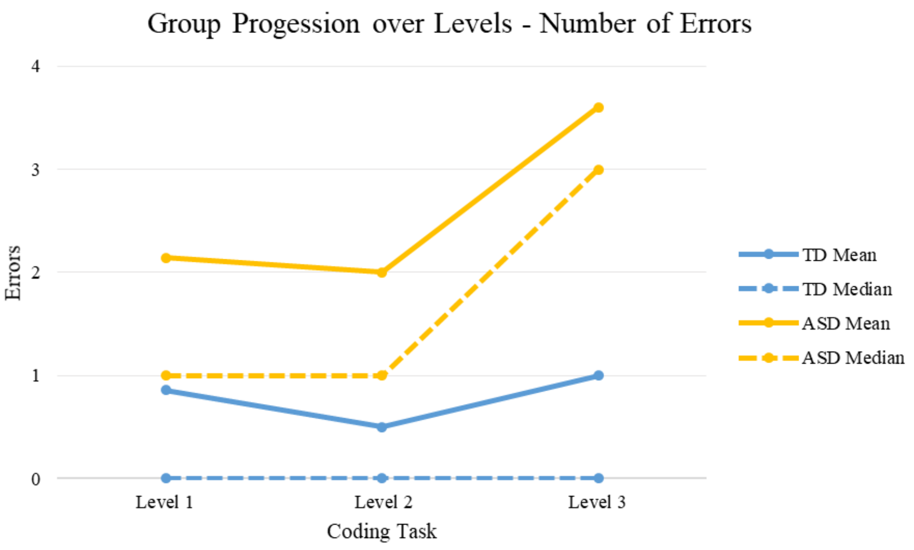


Figure 4.2: Group progression over levels for number of errors in Coding task.

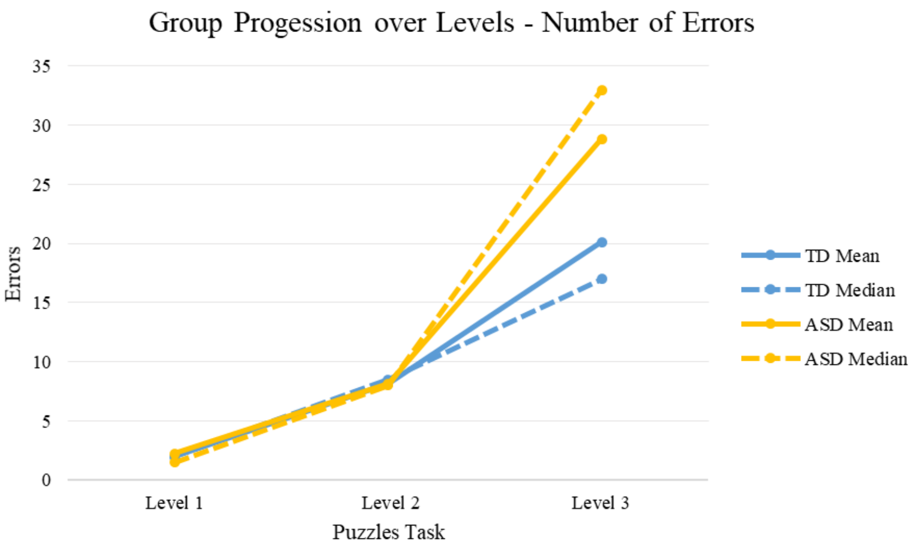


Figure 4.3: Group progression over levels for number of errors in Puzzles task.

Latency time progression per group for the Coding task is represented in Figure 4.4. The mean and median for the experimental group are quite different, which indicates the mean was influenced by an outlier. Therefore, it is better to compare the median curves. The latency time in ASD group increased almost linearly while the TD group latency time was lower in level 3 than level 2. In level 3 the highest difference between groups was observed, however there are also clear differences in levels 1 and 2. In levels 1 and 3 the experimental group had a higher latency time and in level 2 the control group latency time was higher.

For the Puzzles task, the ASD mean and median were also very different indicating once again the outlier influence on the mean (see Figure 4.5). The progression lines were similar for both groups and the differences observed are not as clear as in the Coding task.

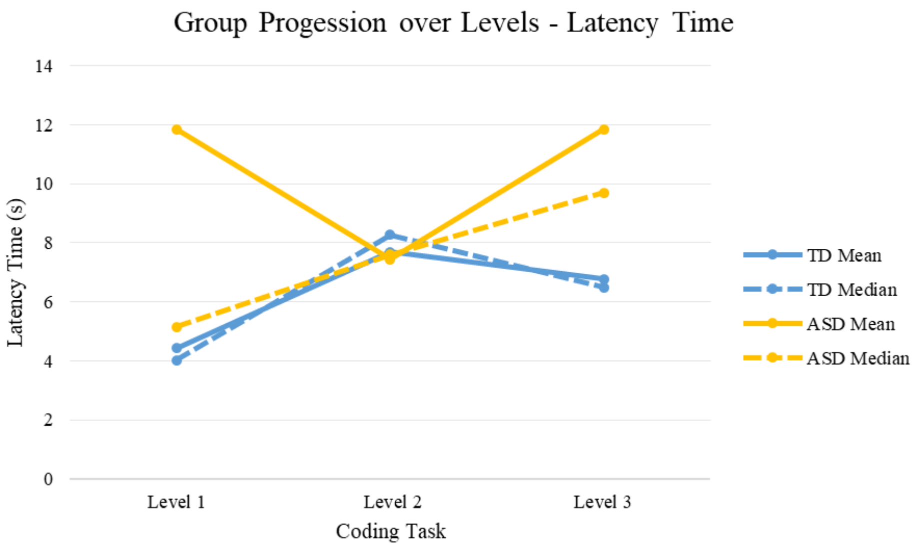


Figure 4.4: Group progression over levels for latency time in Coding task.

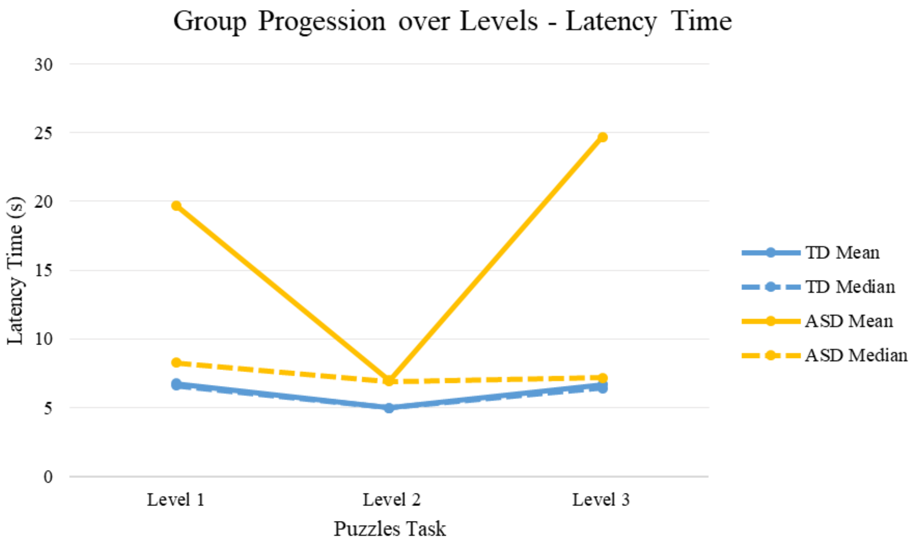


Figure 4.5: Group progression over levels for latency time in Puzzles task.

Finally, total time of execution progression per group for the Coding task is represented in Figure 4.6. The mean and median for the control group are quite different, which indicates the mean was influenced by outliers. The total time in both groups increased almost linearly and was always higher for the ASD group. As verified by the statistical tests, there are significant differences between ASD and TD participants.

For the Puzzles task, the means and medians were very close indicating no outlier influence (Figure 4.7). The progression lines were similar for both groups and the differences observed are not as large or clear as in the Coding task. In levels 1 and 3 the experimental group had higher total time while in level 2, although the difference is very small, the TD group time was higher. The only level with noticeable differences was level 3.

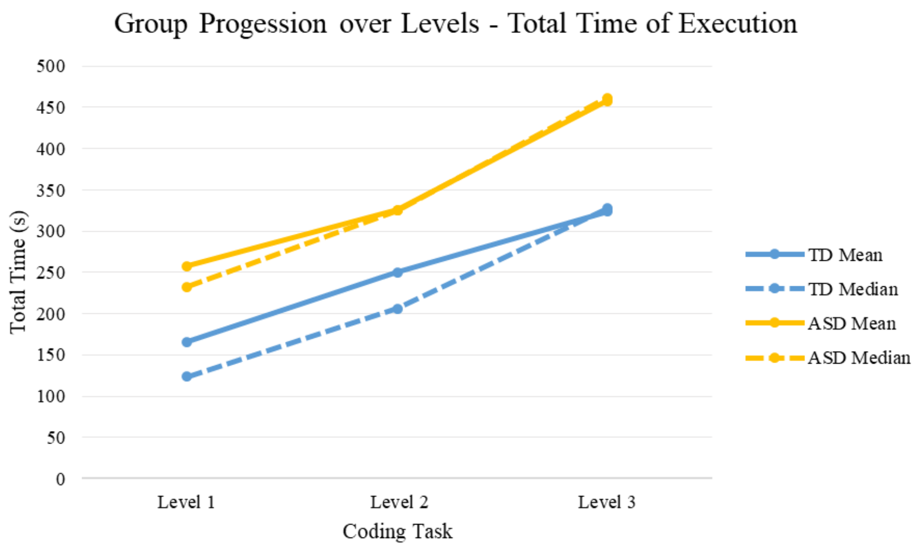


Figure 4.6: Group progression over levels for total time of execution in Coding task.

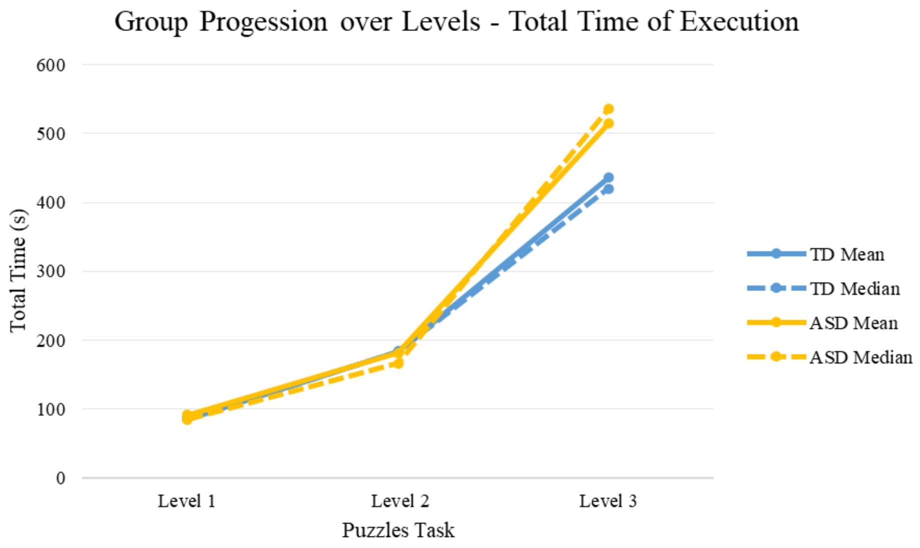


Figure 4.7: Group progression over levels for total time of execution in Puzzles task.

In order to find out if the measures are related, the Spearman correlation factor was calculated for each scene of each group. The results are presented in Table 4.8. The coefficients obtained are very different between groups and tasks. Thus, not allowing to verify any kind of correlation between the measures of interest.

Table 4.8: Spearman ρ for TD and ASD groups in the three levels of Coding and Puzzles tasks.

		TD			ASD		
		E vs. LT	E vs. TT	LT vs. TT	E vs. LT	E vs. TT	LT vs. TT
Coding	Scene 1	0.72	0.80	0.63	-0.15	0.46	0.93
	Scene 2	0.34	0.58	0.72	-0.64	0.13	0.43
	Scene 3	0.09	0.45	0.34	-0.45	0.22	0.70
Puzzles	Scene 1	0.52	0.40	0.03	-0.11	0.18	0.91
	Scene 2	-0.11	0.68	0.22	-0.32	0.23	0.11
	Scene 3	0.09	0.72	0.20	-0.29	0.18	0.29

E – number of errors; LT – latency time; TT – total time of execution

The final group comparison performed was the average time per letter/ piece (Coding/ Puzzles) over scenes. The results obtained are summarized in Table 4.9. As observed, overall average time per letter/ piece is higher for the ASD group than for the TD group, except in Puzzles task level 2.

Table 4.9: Average time per letter/ piece (Coding/ Puzzles task) in seconds for both groups.

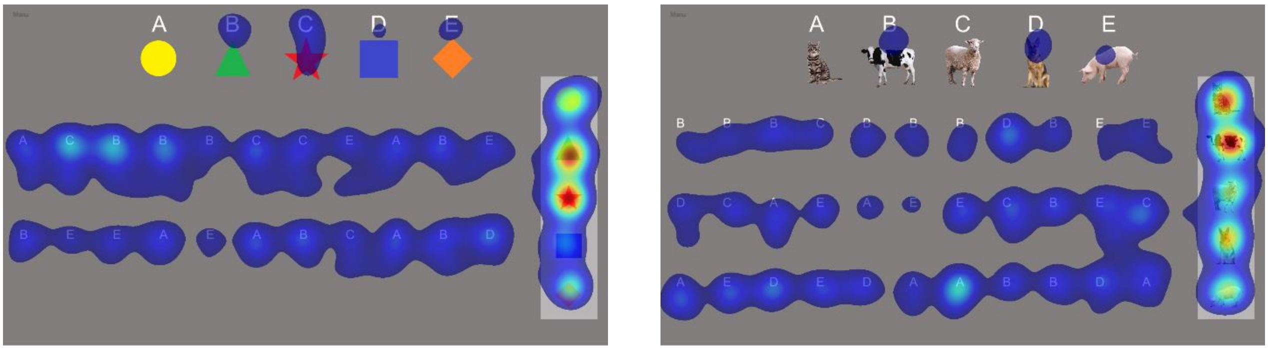
Task	Level (number of letters/ pieces)	Time per Letter/ Piece (s)	
		TD	ASD
Coding	Level 1 (22 letters)	7.5	11.7
	Level 2 (33 letters)	7.6	9.9
	Level 3 (44 letters)	7.4	10.4
Puzzles	Level 1 (6 pieces)	14.2	15.3
	Level 2 (12 pieces)	15.3	15.1
	Level 3 (24 pieces)	18.2	21.5

4.2 GAZE RESULTS

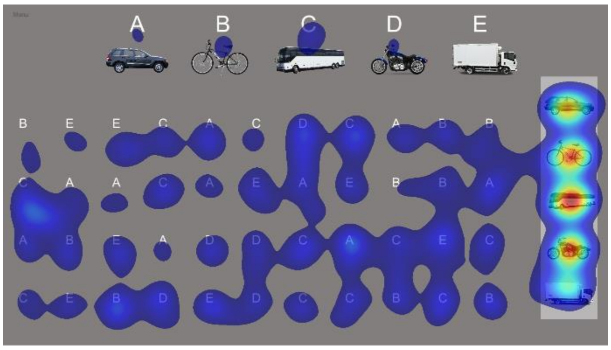
In this subchapter the results from gaze data analysis are presented. The heatmaps shown are from representative participants of each group and the remaining maps are presented in Annex E). Different colours indicate different attention levels to the areas in the scene as demonstrated in Figure 4.8. As these maps were created based on the gaze point coordinates, the areas in red indicate more attention given to that specific area as more time was spent looking there (more points = more time).

Figure 4.8: Colour scale for heatmaps.

Figure 4.9 shows the heatmaps produced for each level of the Coding task of a representative TD participant. The focus in each scene is the sequences and the figure inventory on the right side. There is practically no focus on the code presented on the top of the screen and the attention pattern visualising the sequences is quite similar. The participant looked at every letter in order to complete the level.



a) Scene 1 – Geometric Figures b) Scene 2 – Farm Animals

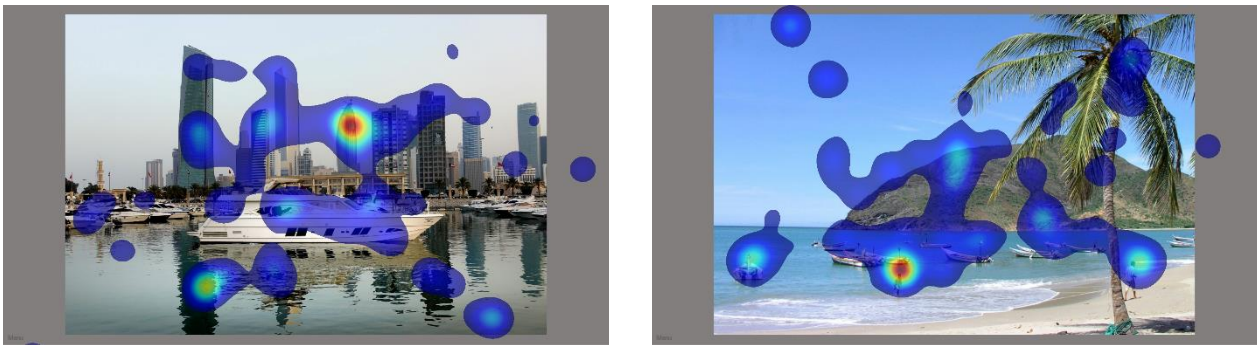


c) Scene 3 – Land Vehicles

Figure 4.9: Examples of heatmaps from representative TD participant in Coding scenes.

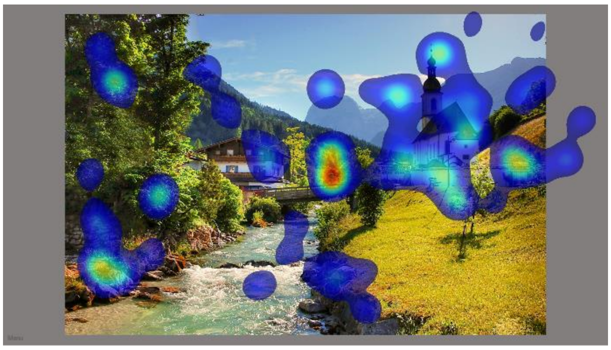
Figure 4.10 shows the heatmaps produced for each presented image in the Puzzles task of a representative TD participant.

In the first picture (Figure 4.10 – a) the participant looked all over the image, focusing on three different elements particularly: background buildings, water and boat. Likewise, in level two (Figure 4.10 – b), the child focused on the main elements in the picture (boats and palm tree) but also looked at the remaining areas to be able to build the puzzle. The last image was the most complex, with more colours and elements. In this case, the participant observed almost every area in the picture, focusing on the central area where there are two reference elements (church and house – Figure 4.10 – c).



a) Scene 1 – City

b) Scene 2 – Beach/ Island

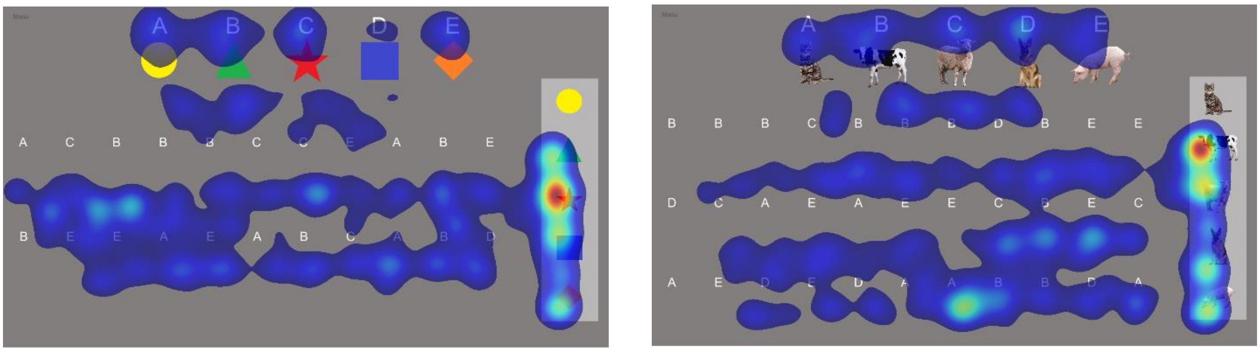


c) Scene 3 – Mountains

Figure 4.10: Examples of heatmaps from representative TD participant in Puzzles scenes.

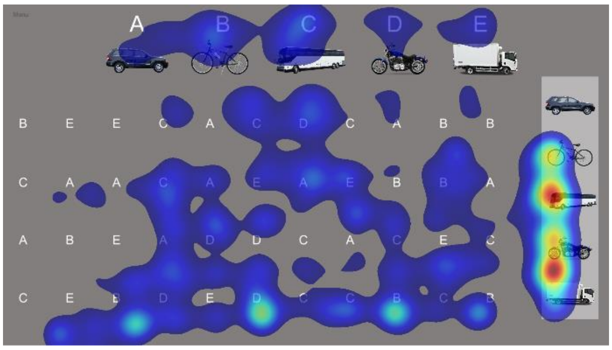
Heatmaps were also created for the ASD participants gaze analysis. The results obtained for the Coding scenes are shown in Figure 4.11.

In the three scenes, the participant focused mostly on the last four images of the inventory and the code presented on top of the screen. The attention pattern in the letter sequences was very uneven in all levels, indicating the child did not look at every letter individually for approximately the same time, but instead looked at the areas bellow the letters, where he had to place the corresponding images. There are also areas of more focus in the last row, in levels two and three (Figure 4.11 – b) and c)) which can indicate some fatigue or distraction in completing the last elements of the task.



a) Scene 1 – Geometric Figures

b) Scene 2 – Farm Animals



c) Scene 3 – Land Transports

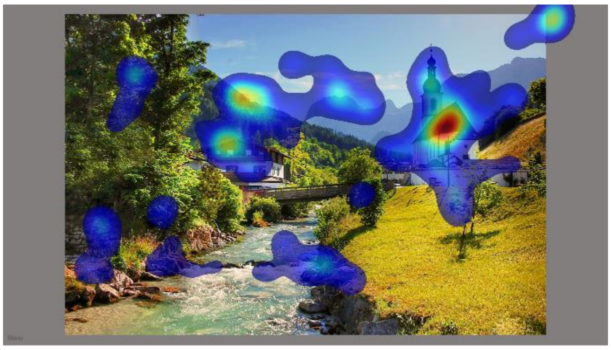
Figure 4.11: Examples of heatmaps from representative ASD participant in Coding scenes.

Finally, Figure 4.12 shows the heatmaps produced for each presented image in the Puzzles task of a representative ASD participant.

In the first picture (Figure 4.12 – a) the participant looked all over the image, focusing on one central element, the boat. In this scene, the child also looked at areas around the picture, revealing a little bit of distraction from the task. In level two (Figure 4.12 – b), the focus was on the main elements in the picture (mainly the palm tree and also one of the boats). In the final level, the participant observed almost every area in the picture, focusing on the church and the house (Figure 4.12 – c).

a) Scene 1 – City

b) Scene 2 – Beach/ Island



c) Scene 3 – Mountains

Figure 4.12: Examples of heatmaps from representative ASD participant in Puzzles scenes.

The heatmaps for all the remaining participants are presented in Annex E).

5 DISCUSSION

This chapter is dedicated to the discussion of the results presented in Chapter 3.

Before analysing the score and gaze results, a group comparison for age was done. This was an important step in order to understand if the results obtained could be influenced by significant age differences between the participants in both groups. According to Figure 4.1, the age median for the two groups was 8 years, however the distribution of ages is different between groups. Most participants in the control group have between 7 and 8 years, while the participants from the experimental group have between 6 and 9 years. Although the distributions are different, these differences are not significant, indicating the results obtained in the statistical tests are not significantly influenced by the participants' ages. This result indicates that the groups are homogenous regarding this demographic variable and therefore comparable. The same result was verified for gender but as mentioned before, there were differences in both groups (gender) that could influence the results and should be addressed in future work.

5.1 SCORE RESULTS

From a general point-of-view, the ASD participants showed a tendency to worst overall performance on both tasks than the TD participants. As expected, based on previous studies, most significant differences were verified for the Coding task's total time of execution.

Regarding the Coding task, the results obtained from the Mann-Whitney U tests showed significant differences for the total time of execution, with the ASD group having a higher time in all three levels. According to Table 4.7, these differences were higher in level 3 (p – value = 0.009) than in the first two levels (p – value = 0.015 in level 1 and p – value = 0.016 in level 2).

Another way to visualise the progression over the three levels is observing the progression graphics. Figure 4.6 shows the groups' progression over levels for total time of execution. Both groups had a fairly linear progression over levels, but in level 3, the ASD participants had a higher increase in total time compared to the TD participants. This increase may be an indicator of the children's difficulty to keep the same levels of attention and process more stimuli while keeping an adequate speed of execution. In other words, with time ASD children tend to execute the task more slowly. These results and the task completion percentage for both groups (Table 4.5: TD = 100%, ASD = 62.5%) may indicate that ASD children not only have worse performance in this task but also, that they may not be able tolerate the frustration associated with it. A simple way to observe these difficulties is to compare the number of errors.

Although the permutation tests did not show significant differences (Table 4.6: p – values > 0.05), by comparing the results in Tables 4.1 and 4.3, it is clear that the ASD participants had more difficulties executing the Coding task. For the TD group, the median of the number of errors per scene was 0 and the maximum was 6, 4 and 5, for levels 1, 2 and 3, respectively. For the ASD group these values were always higher (median per scene: 1, 1, 3; maximum: 8, 8 and 9). In each level, the total number of errors was always higher in the ASD group, as demonstrated in Figure 4.2. The progression over levels for this measure was linear for the TD group, however for the

ASD group there was a large increase in number of errors in level 3. These may results indicate difficulties in completing long duration tasks with a larger number of stimuli.

The results concerning the latency time in the Coding task were not as different as for the other two measures, as demonstrated in Figure 4.4. TD participants had a lower latency time in levels 1 and 3, however in level 2 the median latency time was higher than for the ASD participants. This group had more latency time in each level compared to the previous but in level 2 it was smaller than the TD group. Though there are differences between both groups concerning latency time, the Mann-Whitney U tests did not show any significant differences (Table 4.7: *p* – values > 0.05). The tendency to have non-significant differences may be explained by the sample size. A bigger sample could show that these differences are after all significant. Although harder to understand, the latency time results showed a pattern. In this task it increases in level 2 and decreases in level 3. This is possibly explained by the fact that in level 2 there is a stimulus change – geometric figures to farm animals – which cognitively may implicate longer processing since the child needs to process color, shape, content (what animals are these). In scene 3, latency time reduces because the type of stimuli presented is similar to the previous scene.

Looking at the progression graphics for the Coding task (Figures 4.2, 4.4 and 4.6), it is possible to see the group evolution is very similar in two of the three measures of interest – number of errors and total time of execution. However, it was not possible to establish a correlation between the measures of interest as demonstrated in Table 4.8. By looking at Table 4.8 there is a difference between the two groups in the correlation between number of errors and total time of execution – in general, it is possible to observe strong positive correlations in the TD group but not in the ASD group. Concerning the correlation between latency time and total time, it is stronger and more evident in the ASD group, especially in level 1 of both tasks. The correlation between latency time and number of errors is harder to understand. However, there is a tendency for negative correlation in the ASD group but not in the TD group.

Considering the results obtained for the three measures of interest and the completion percentage, it is possible to conclude the ASD group had more difficulties executing long duration tasks with a larger number of stimuli, leading to a worse overall performance. From an intervention perspective this is a very interesting conclusion. By observing the heatmaps in Figure 4.11, it is clear that ASD children focus in the central areas of the screen, ignoring stimuli in peripheral areas. Also, the gaze is less precise with a lot of movement from one point to another. These tendencies are also visible in Figure 4.12, for the Puzzles task. This information is very important for intervention because it could indicate that ASD children attention focus is smaller therefore intervention materials should have stimuli more concentrated in the central region of the page/ screen.

Regarding the Puzzles task, the results from the statistical tests did not show any significant differences for any measure of interest as Tables 4.6 and 4.7 indicate (*p* – values > 0.05). In this task the participants' performance was more similar, however it is still possible to observe some difference trend in performance between groups.

Looking at Tables 4.2 and 4.4, it is possible to verify the experimental group had more errors than the control group only in level 3, considering the median values (TD = 2.00, 8.50, 17.00 and ASD = 1.50, 8.00, 33.00). As mentioned before, the mean is very sensitive to outliers, so it is more appropriate to consider the median. However, the maximum number of errors per scene was higher for the ASD group in all levels (TD = 5, 17, 40; ASD = 7, 21, 41). According to the permutation test results, these differences are not significant (Table 4.6: *p* – value > 0.05), except for level 3 (*p* – value = 0.039). A better way to visualise group differences is looking at the progression graphic in Figure 4.3. TD participants had a linear progression over levels, whereas ASD participants had a larger increase in level 3.

The ASD participants' latency time in the Puzzles task was higher than the TD participants' time. However, the differences are not significant according to the Mann-Whitney U tests (Table 4.7: p – value > 0.05). In Figure 4.5 the differences between groups is small but visible. The groups had a similar progression over levels and level 2 was the one where the latency time was lower in both groups. In this task, the latency time results had a pattern – higher in levels 1 and 3 and lower in level 2. A possible justification for this pattern is the fact that the picture selected for level 2 is easier to memorize since it has less elements and central elements that facilitate the memorization.

Mann-Whitney U tests did not show significant differences for total time of execution in the Puzzles task (Table 4.7: p – value > 0.05) and the progression over levels is very similar for both groups (Figure 4.7). The TD group progression was almost linear whereas the ASD group had a much larger increase in total time of execution in level 3. Once again, this larger increase may be an indicator of the children's inability to complete the task when the number of pieces is larger. These results and the task completion percentage for both groups (Table 4.5: TD = 100%, ASD = 87.5%) may indicate that ASD children have worse performance in the task and also, that they are not able tolerate the frustration associated with it. It is noteworthy that even though the completion percentage was still smaller in the ASD than the TD, it was higher in the Puzzles task which indicates that ASD children perform better in this task than in Coding and, consequently, prefer it, being able to tolerate the frustration generated by the errors and uncertainty better.

The final group comparison done was the calculation of the average time per letter or piece (Table 4.9). These results showed once again the ASD children had more difficulties than the TD children, especially in the Coding task where the differences were larger (TD (s) = 7.5, 7.6, 7.4; ASD (s) = 11.7, 9.9, 10.4). Even comparing the average time per letter/piece between levels, the TD participants had a very similar time over levels, whereas the ASD participants had more difficulties in the first level. This indicates that in the first level children were still unfamiliar with the task and had to spend more time per letter. In the following levels, the time per piece was shorter, suggesting that they understood the task and improved their performance with practice. In the Puzzles task there was always a higher time per piece at each level (TD (s) = 14.2, 15.3, 18.2; ASD (s) = 15.3, 15.1, 21.5) suggesting the increase in difficulty made the children take more time to place each piece.

Considering the score results obtained during the testing phase there are some conclusions to take. In terms of number of errors, it is normal to have more errors at each level because the difficulty increases (more letters or pieces), yet in the Coding task this was not observed for the TD group and for the ASD group it was only observed in level 3 (Figure 4.2). This shows that the participants were able to understand the task and complete it without a higher number of errors, despite the difficulty increase in each scene. In the Puzzles task, the number of errors increases in each level (Figure 4.3) which means, in this case, the increasing difficulty did affect the ability to complete the level without making more errors. Combining these results with the average time per letter/piece it is possible to conclude that in the Coding task, the participants were able to progress in each level despite the increasing difficulty, and with practice it is possible to improve the overall performance.

In relation to total time of execution, in both tasks, there is a larger increase in level 3 for the two groups. Again, this may be an indicator of the children's difficulty to keep the same levels of attention and rebuild an image divided in more pieces, while keeping an adequate speed of execution. Considering the survey results in Annex B), 54.2% of the participants preferred the Puzzles task and 79.2% considered this was the hardest task, meaning that most children preferred the hardest task. Therefore, increased time of execution in the final levels of each task is probably related to the difficulties in completing a task with a higher number of stimuli. This was particularly visible in ASD participants because they needed a lot more encouragement.

Finally, it is important to mention that 95.8% of the children who tried the game said they liked it (Annex B) and despite some difficulties, 73.1% of the participants finished all the levels in the game. These results demonstrated a very good acceptance of the game.

5.2 GAZE RESULTS

The gaze analysis was very important in order to understand the differences between groups regarding how children explored visually the game and possibly explain some of the score results obtained. It is important to remember the results presented in Chapter 4.2 are from two representative participants (one from each group).

Regarding the Coding task, there are clear differences between the two groups. The TD group almost did not look at the code to complete the task, mainly focusing on the inventory images (Figure 4.9). In contrast, the ASD group needed to look much more frequently at the code on the top of the screen to complete the task and practically did not look at the first image in the inventory (Figure 4.11). Another relevant difference is the progression over levels. TD children looked at the code less frequently as the game progressed, while ASD children kept on looking at the code throughout the game's levels. This suggests that TD children were more capable of memorizing the code given, having less need to check the code in each level and a better working memory. ASD children did progress in terms of average time per piece (Chapter 5.1) but were not able to complete the scene without confirming the correspondence. This need to look at the code is another explanation for the higher time of execution of this group and a possible indicator of lower working memory.

Observing the heatmaps for the Coding task there is also a difference in the attention pattern. TD participants looked at all the letters in the sequences evenly. This means they were more organized in terms of gaze focus and therefore were able to complete the task in less time. In comparison, ASD participants had a more disorganized gaze and did not focus equally in every letter in the sequences, contributing once more to a slower performance.

Concerning the Puzzles task, the heatmaps obtained are also quite different between groups. The TD group focused especially on the main elements of each picture but looked at all the different areas to memorise most of the image (Figure 4.10). Comparatively, the ASD group focused especially in one element of each picture and just a little on some of the remaining areas (Figure 4.12). In addition to that, in scene 1, participants also focused on the borders around the picture indicating lack of concentration in the task. These factors contribute for a worse performance in completing the task and can also indicate that ASD children may have better memory and don’t feel the need to look at the image as thoroughly since the differences between groups regarding number of errors, latency times and total times of execution were not significant (discussed in Chapter 5.1).

6 CONCLUSION

During the course of this project, a game was created, based on the currently used tasks utilised to evaluate and help the evaluation of executive profile in ASD in children. The developed game was well accepted by the participants which suggests that it could be further used and developed into a more complete medical software to be used by clinical psychologists.

The results that were obtained using the eye tracking device are mostly consistent with what was expected *a priori*: TD children are more capable in resolving the tasks of the game, showing a faster total time of execution and a smaller amount of errors when compared with ASD children. It is possible to conclude – from the analysis of the total time of execution, latency time and number of errors – that the ASD group had more difficulties understanding and executing the game tasks, leading to a worse overall performance; even though the permutation test and Mann-Whitney U test applied did not show significant differences in the three measures, which we consider to be a consequence of a small sample of ASD children.

As it came to be expected from the results of previous studies, the most significant differences between groups occurred in the “Coding” task of the game.

In the “Puzzle” task, there were no significant differences in the three aforementioned measures, according to the statistical tests, although the ASD group registered a longer latency time. The progression over the levels is similar between the groups and one can note the shorter latency time for both groups in the second level of this task, which showed a relatively simple image of a beach and so required less time to remember before starting the image assembly.

From an intervention perspective it would be important for ASD children to practice the processing of a higher number of simultaneous stimuli, since they showed a worse performance in long duration tasks (level 3 of each task) with more stimuli.

With respect to the heatmaps, the comparison between TD children and ASD children shows clear differences in the way each group uses eye movements to obtain information before and during the tasks. TD children show a more organised approach in terms of gaze focus, focusing equally on every letter during the “Coding” task, while ASD children show a more disorganized gaze focus and seem to favour some letters over others. Furthermore, TD children show improved working memory by looking less frequently at the code on the top of the screen as the game progresses, while ASD children keep looking at the code with almost the same frequency throughout the game. The differences between groups keep showing themselves in the “Puzzles” task: while TD children focused on the picture’s main area, they still moved their eyes across it in its entirety to memorise more details; ASD children seem to focus on a single detail of a picture and mostly ignore the rest. These differences in eye movements between TD and ASD children are evident and easy to identify and so could be used as a positive sign of ASD when testing using this method. These gaze results can provide very important information for intervention in ASD since they clearly demonstrate a reduced visual area of attention. Thus, all the stimuli presented in intervention tools should be more centralized.

6.1 LIMITATIONS

The limitations faced by this study will now be discussed. The Tobii eye tracker license used was the one available for free; while this version of software allowed the acquisition of gaze data and the integration into Unity in a fairly simple manner, some advanced features that would have been useful for further analysis – like the fixation calculation algorithm – were not available without the paid license.

Furthermore, the error counting method was limited in the sense that it would count errors that should not be counted as such, which was considered when analysing the data but should be improved upon in future work.

The sample size of both groups was also different because, during the short data collection period, it was not possible to find more ASD children that fit the precise requirements of this study and whose parents or guardians were willing to have them take part in the study; nevertheless, having a larger sample size for both groups and performing more tests per participant (to evaluate the effect repeated contact with the tasks) would have yielded a more robust final result.

Finally, using the mouse as the input device could be considered a limitation because not all children had had previous contact with a mouse and so had not developed a minimum of dexterity in moving it and using it to select the various objects. If the game had a touch screen interface, perhaps some children would have had an easier time, since nowadays most children have access to and enjoy playing on their parents' or guardians' smartphones and tablet computers.

6.2 FUTURE WORK

As for future work, as mentioned before, the game itself could be improved in order to present different challenges and test other skills that could be relevant to characterize executive profile in ASD. Further than that, any future testing should strive to have more participants and do more sessions with each participant, in order to evaluate each participant's progress over time, if there is any.

Apart from that, the study has shown clear differences in performance between ASD and TD children and so stands to be used as groundwork for the development of a new and cheaper method for evaluation of the executive profile and possible intervention in ASD for children between 6 and 10 years of age.

Finally, it would be interesting to compile a mobile version for Android/ iOS even without the eye tracking component in order to evaluate the app's potential as an intervention tool for parents/ caretakers and in clinical environment.

BIBLIOGRAPHY

[1] J. D. Enderle and J. D. Bronzino, *Introduction to Biomedical Engineering*. Elsevier Inc., 2011.

[2] “Biomedical Engineers : Occupational Outlook Handbook: U.S. B.” [Online]. Available: https://www.bls.gov/ooh/architecture-and-engineering/biomedical-engineers.htm#tab-6. [Accessed: 12-Aug-2019].

[3] Tobii, “Tobii Eye Tracking - An introduction to eye tracking and Tobii Eye Trackers,” *Technology*, p. 12, 2010.

[4] “Tobii Pro Learn & Support,” 2015. [Online]. Available: https://www.tobiipro.com/learn-and-support/. [Accessed: 09-Aug-2019].

[5] “How eye tracking works - Tobii Dynavox.” [Online]. Available: https://www.tobiidynavox.com/en-us/about/about-us/how-eye-tracking-works/. [Accessed: 09-May-2019].

[6] A. M. Feit *et al.*, “Toward everyday gaze input: Accuracy and precision of eye tracking and implications for design,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2017, vol. 2017-May, pp. 1118–1130, doi: 10.1145/3025453.3025599.

[7] Tobii Technology, “Accuracy and precision test method for remote eye trackers,” *Test*, pp. 1–28, 2011.

[8] “HeatMapCo Instant Eye-Tracking Prediction Website Heat Map.” [Online]. Available: http://heat-map.co/. [Accessed: 27-Jun-2019].

[9] “Heatmaps.” [Online]. Available: https://www.optimizely.com/optimization-glossary/heatmap/. [Accessed: 27-Jun-2019].

[10] D. Purves *et al.*, *Types of Eye Movements and Their Functions*. Sinauer Associates, 2001.

[11] “APA Dictionary of Psychology.” [Online]. Available: https://dictionary.apa.org/. [Accessed: 14-Aug-2019].

[12] A. Diamond, “Executive functions.,” *Annu. Rev. Psychol.*, vol. 64, pp. 135–68, 2013, doi: 10.1146/annurev-psych-113011-143750.

[13] A. B. Sereno and M. S. Bolding, “Executive Functions: Eye Movements and Human Neurological Disorders,” *Ref. Modul. Neurosci. Biobehav. Psychol.*, Jan. 2017, doi: 10.1016/B978-0-12-809324-5.02099-X.

[14] M. Solomon, S. J. Ozonoff, N. Cummings, and C. S. Carter, “Cognitive control in autism spectrum disorders.,” *Int. J. Dev. Neurosci.*, vol. 26, no. 2, pp. 239–47, Apr. 2008, doi: 10.1016/j.ijdevneu.2007.11.001.

[15] M. T. Carlin, S. Soraci, A. L. Goldman, and W. McIlvane, “Visual search in unidimensional arrays: A comparison between subjects with and without mental retardation,” *Intelligence*, vol. 21, no. 2, pp. 175–196, Sep. 1995, doi: 10.1016/0160-2896(95)90025-X.

[16] E. Vakil, H. Lifshitz, D. Tzuriel, I. Weiss, and Y. Arzuoan, “Analogies solving by individuals with and without intellectual disability: Different cognitive patterns as indicated by eye movements,” *Res. Dev. Disabil.*, vol. 32, no. 2, pp. 846–856, Mar. 2011, doi: 10.1016/j.ridd.2010.08.006.

[17] K. Rayner, “Eye Movements and Cognitive Processes in Reading, Visual Search, and Scene Perception,” *Stud. Vis. Inf. Process.*, vol. 6, pp. 3–22, Jan. 1995, doi: 10.1016/S0926-907X(05)80003-0.

[18] G. F. Eden, J. F. Stein, H. M. Wood, and F. B. Wood, “Differences in eye movements and reading problems in dyslexic and normal children,” *Vision Res.*, vol. 34, no. 10, pp. 1345–1358, May 1994, doi: 10.1016/0042-6989(94)90209-7.

[19] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, “Visual Fixation Patterns During Viewing of Naturalistic Social Situations as Predictors of Social Competence in Individuals With Autism,” *Arch. Gen. Psychiatry*, vol. 59, no. 9, p. 809, Sep. 2002, doi: 10.1001/archpsyc.59.9.809.

[20] C. von Hofsten, H. Uhlig, M. Adell, and O. Kochukhova, “How children with autism look

at events," *Res. Autism Spectr. Disord.*, vol. 3, no. 2, pp. 556–569, Apr. 2009, doi: 10.1016/J.RASD.2008.12.003.

[21] T. Nakano *et al.*, "Atypical gaze patterns in children and adults with autism spectrum disorders dissociated from developmental changes in gaze behaviour," *Proc. R. Soc. B Biol. Sci.*, vol. 277, no. 1696, pp. 2935–2943, Oct. 2010, doi: 10.1098/rspb.2010.0587.

[22] M. Murias *et al.*, "Validation of eye-tracking measures of social attention as a potential biomarker for autism clinical trials," *Autism Res.*, vol. 11, no. 1, pp. 166–174, Jan. 2018, doi: 10.1002/aur.1894.

[23] A. Moore *et al.*, "The geometric preference subtype in ASD: identifying a consistent, early-emerging phenomenon through eye tracking.," *Mol. Autism*, vol. 9, p. 19, 2018, doi: 10.1186/s13229-018-0202-z.

[24] M. Frost-Karlsson *et al.*, "Social scene perception in autism spectrum disorder: An eye-tracking and pupillometric study," *J. Clin. Exp. Neuropsychol.*, pp. 1–9, Jul. 2019, doi: 10.1080/13803395.2019.1646214.

[25] "Unity Real-Time Development Platform | 3D, 2D VR & AR Visualizations." [Online]. Available: https://unity.com/. [Accessed: 10-Jan-2019].

[26] T. Wright, S. De Ribaupierre, and R. Eagleson, "Design and evaluation of an augmented reality simulator using leap motion," vol. 4, pp. 210–215, 2017, doi: 10.1049/htl.2017.0070.

[27] G. Wheeler *et al.*, "Virtual interaction and visualisation of 3D medical imaging data with VTK and Unity," pp. 1–6, 2018, doi: 10.1049/htl.2018.5064.

[28] C. Bartneck, M. Soucy, K. Fleuret, and E. B. Sandoval, "The Robot Engine - Making The Unity 3D Game Engine Work For HRI," pp. 431–437, 2015.

[29] C. Bandeira de Lima, *Perturbações do espectro do autismo : manual prático de intervenção*. Lidel, 2012.

[30] C. Bandeira de Lima, *Perturbações do neurodesenvolvimento : manual de orientações diagnóstias e estratégias de intervenção*. Lidel, 2015.

[31] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders*. 2013.

[32] E. Billstedt, C. Gillberg, and C. Gillberg, "Autism after adolescence: Population-based 13- to 22-year follow-up study of 120 individuals with autism diagnosed in childhood," *J. Autism Dev. Disord.*, 2005, doi: 10.1007/s10803-005-3302-5.

[33] C. Woolger, "Wechsler Intelligence Scale for Children-Third Edition (wisc-iii)," in *Understanding Psychological Assessment*, Boston, MA: Springer US, 2001, pp. 219–233.

[34] "Wechsler Intelligence Scale for Children (WISC) | Wechsler IQ Test." [Online]. Available: https://wechsleriqtest.com/wechsler-intelligence-scale-for-children/. [Accessed: 30-Aug-2019].

[35] S. Mouga, C. Café, J. Almeida, C. Marques, F. Duque, and G. Oliveira, "Intellectual Profiles in the Autism Spectrum and Other Neurodevelopmental Disorders," *J. Autism Dev. Disord.*, vol. 46, no. 9, pp. 2940–2955, Sep. 2016, doi: 10.1007/s10803-016-2838-x.

[36] "Unity - Manual: Unity User Manual (2019.2)." [Online]. Available: https://docs.unity3d.com/Manual/index.html. [Accessed: 05-Dec-2018].

[37] "PNG images with transparent background | Free PNG images clipart." [Online]. Available: http://pngimg.com/. [Accessed: 07-Jan-2019].

[38] "Free Sound Clips | SoundBible.com." [Online]. Available: http://soundbible.com/. [Accessed: 05-Mar-2019].

[39] "Tobii Pro Upgrade Key for Tobii Eye Tracker 4C." [Online]. Available: https://www.tobiipro.com/siteassets/tobii-pro/product-descriptions/tobii-pro-upgrade-key-product-description.pdf/. [Accessed: 12-Dec-2018].

[40] "EUR-Lex - 02016R0679-20160504 - EN - EUR-Lex." [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1532348683434&uri=CELEX:02016R0679-20160504. [Accessed: 28-May-2019].

[41] "R: The R Project for Statistical Computing." [Online]. Available: https://www.r-project.org/. [Accessed: 15-Jul-2019].

[42] "GitHub - TobiasRoeddiger/GazePointHeatMap 👁." [Online]. Available:

https://github.com/TobiasRoeddiger/GazePointHeatMap. [Accessed: 29-Jul-2019].

[43] H. Wickham and L. Stryjewski, “40 Years of Boxplots,” *Had.Co.Nz*, pp. 1–17, 2011, doi: 10.1126/science.225.4661.519.

[44] Alan Agresti, “A Survey of Exact Inference for Contingency Tables,” *Stat. Sci.*, vol. 7, no. No. 1, pp. 131–153, 1992.

[45] H. A. David, “The Beginnings of Randomization Tests,” *Am. Stat.*, vol. 62, no. 1, pp. 70–72, Feb. 2008, doi: 10.1198/000313008X269576.

[46] T. Hothorn, K. Hornik, and M. A. Van De Wiel, “Implementing a Class of Permutation Tests : The coin Package.”

[47] M. P. Fay and M. A. Proschan, “Wilcoxon-Mann-Whitney or t-test? On assumptions for hypothesis tests and multiple interpretations of decision rules.,” *Stat. Surv.*, vol. 4, pp. 1–39, 2010, doi: 10.1214/09-SS051.

ANNEXES

A) RANDOM SEQUENCES

	Linha 1		Linha 2		Linha 3		Linha 4	
Cena 1	1	A	2	B	4	D	5	E
	3	C	5	E	5	E	1	A
	2	B	5	E	2	B	4	D
	2	B	1	A	4	D	4	D
	2	B	5	E	4	D	3	C
	3	C	1	A	2	B	4	D
	3	C	2	B	1	A	4	D
	5	E	3	C	1	A	3	C
	1	A	1	A	2	B	2	B
	2	B	2	B	4	D	5	E
	5	E	4	D	4	D	4	D

a)

	Linha 1		Linha 2		Linha 3		Linha 4	
Cena 2	2	B	4	D	1	A	4	D
	2	B	3	C	5	E	5	E
	2	B	1	A	4	D	2	B
	3	C	5	E	5	E	1	A
	2	B	1	A	4	D	5	E
	2	B	5	E	1	A	3	C
	2	B	5	E	1	A	1	A
	4	D	3	C	2	B	5	E
	2	B	2	B	2	B	1	A
	5	E	5	E	4	D	3	C
	5	E	3	C	1	A	2	B

b)

	Linha 1		Linha 2		Linha 3		Linha 4	
Cena 3	2	B	3	C	1	A	3	C
	5	E	1	A	2	B	5	E
	5	E	1	A	5	E	2	B
	3	C	3	C	1	A	4	D
	1	A	1	A	4	D	5	E
	3	C	5	E	4	D	4	D
	4	D	1	A	3	C	3	C
	3	C	5	E	1	A	3	C
	1	A	2	B	3	C	2	B
	2	B	2	B	5	E	3	C
	2	B	1	A	3	C	2	B

c)

Figure 3.13: Random sequences generated to define letters order in Coding task (level 1 – a), level 2 – b), level 3 – c)).

Table 3.3: Puzzles task level 1 - presentation order based on random sequence.

	Piece Number	Random Number	Piece Name
Scene 5 BD	5	0.306	B2
	4	0.105	B1
	1	0.836	A1
	2	0.105	A2
	6	0.946	B3
	3	0.641	A3

Table 3.4: Puzzles task level 2 - presentation order based on random sequence.

	Piece Number	Random Number	Piece Name
Scene 7 BD	9	0.661	C1
	2	0.754	A2
	8	0.926	B4
	6	0.533	B2
	3	0.230	A3
	10	0.893	C2
	1	0.566	A1
	7	0.804	B3
	5	0.929	B1
	12	0.100	C4
	11	0.720	C3
	4	0.012	A4

Table 3.5: Puzzles task level 3 - presentation order based on random sequence.

	Piece Number	Random Number	Piece Name
Scene 9 BD	19	0.331	D1
	7	0.831	B1
	11	0.293	B5
	3	0.767	A3
	10	0.181	B4
	20	0.190	D2
	6	0.370	A6
	12	0.428	B6
	18	0.128	C6
	5	0.104	A5
	9	0.539	B3
	22	0.300	D4
	24	0.714	D6
	15	0.951	C3
	13	0.122	C1
	1	0.903	A1
	17	0.123	C5
	14	0.542	C2
	23	0.643	D5
	16	0.287	C4
	21	0.140	D3
	2	0.775	A2
	4	0.986	A4
	8	0.994	B2

B) SURVEY RESULTS

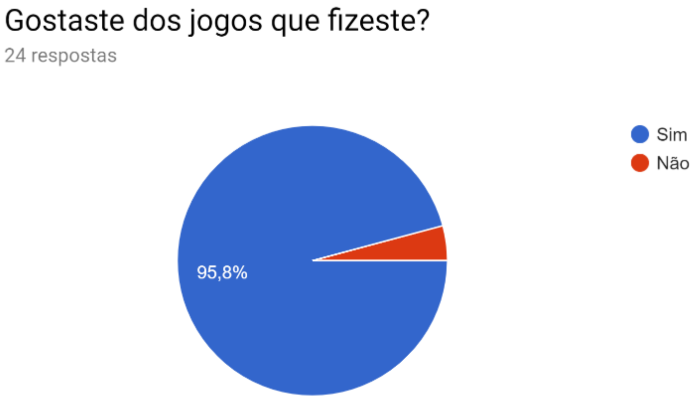


Figure 3.14: Survey Question 1: "Did you like the games played?"

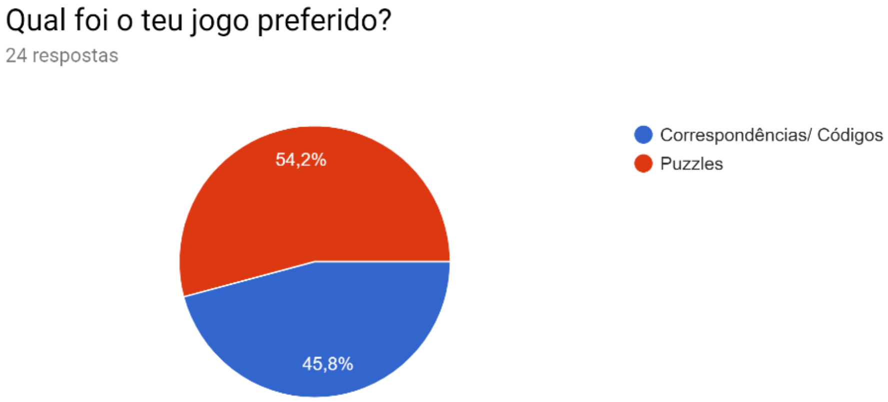


Figure 3.15: Survey Question 2 - "What was your favorite game?"

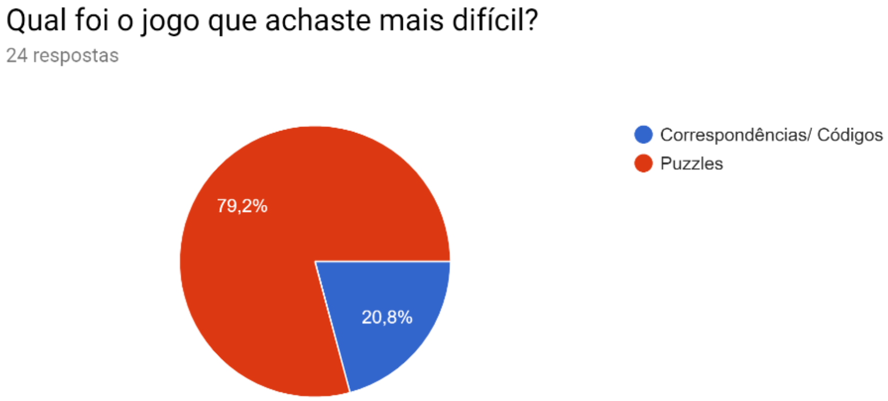
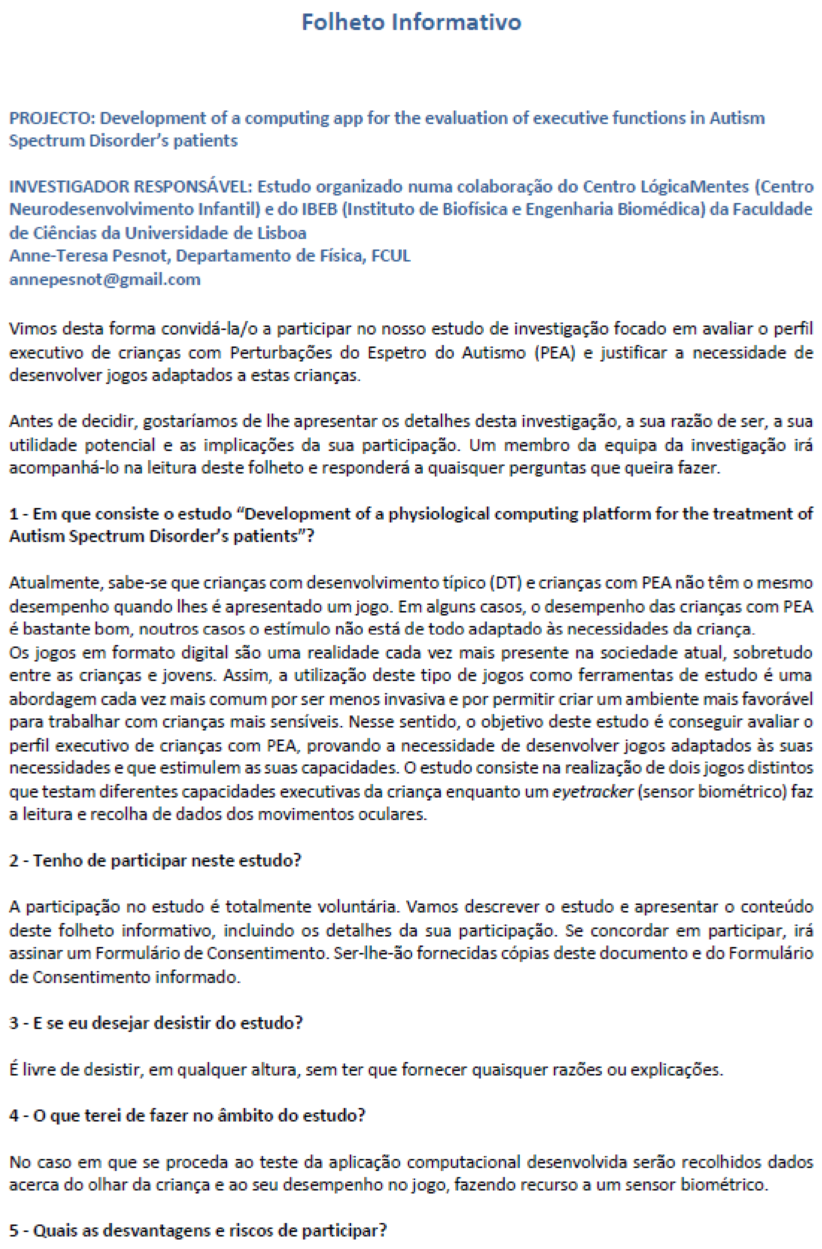


Figure 3.16: Survey Question 3 - "Which game did you find harder?"

C) INFORMATIVE LEAFLET AND INFORMED CONSENT



Folheto Informativo

PROJECTO: Development of a computing app for the evaluation of executive functions in Autism Spectrum Disorder's patients

INVESTIGADOR RESPONSÁVEL: Estudo organizado numa colaboração do Centro LógicaMentes (Centro Neurodesenvolvimento Infantil) e do IBEB (Instituto de Biofísica e Engenharia Biomédica) da Faculdade de Ciências da Universidade de Lisboa
Anne-Teresa Pesnot, Departamento de Física, FCUL
annepesnot@gmail.com

Vimos desta forma convidá-la/o a participar no nosso estudo de investigação focado em avaliar o perfil executivo de crianças com Perturbações do Espetro do Autismo (PEA) e justificar a necessidade de desenvolver jogos adaptados a estas crianças.

Antes de decidir, gostaríamos de lhe apresentar os detalhes desta investigação, a sua razão de ser, a sua utilidade potencial e as implicações da sua participação. Um membro da equipa da investigação irá acompanhá-lo na leitura deste folheto e responderá a quaisquer perguntas que queira fazer.

1 - Em que consiste o estudo "Development of a physiological computing platform for the treatment of Autism Spectrum Disorder's patients"?

Atualmente, sabe-se que crianças com desenvolvimento típico (DT) e crianças com PEA não têm o mesmo desempenho quando lhes é apresentado um jogo. Em alguns casos, o desempenho das crianças com PEA é bastante bom, noutros casos o estímulo não está de todo adaptado às necessidades da criança.
Os jogos em formato digital são uma realidade cada vez mais presente na sociedade atual, sobretudo entre as crianças e jovens. Assim, a utilização deste tipo de jogos como ferramentas de estudo é uma abordagem cada vez mais comum por ser menos invasiva e por permitir criar um ambiente mais favorável para trabalhar com crianças mais sensíveis. Nesse sentido, o objetivo deste estudo é conseguir avaliar o perfil executivo de crianças com PEA, provando a necessidade de desenvolver jogos adaptados às suas necessidades e que estimulem as suas capacidades. O estudo consiste na realização de dois jogos distintos que testam diferentes capacidades executivas da criança enquanto um *eyetracker* (sensor biométrico) faz a leitura e recolha de dados dos movimentos oculares.

2 - Tenho de participar neste estudo?

A participação no estudo é totalmente voluntária. Vamos descrever o estudo e apresentar o conteúdo deste folheto informativo, incluindo os detalhes da sua participação. Se concordar em participar, irá assinar um Formulário de Consentimento. Ser-lhe-ão fornecidas cópias deste documento e do Formulário de Consentimento informado.

3 - E se eu desejar desistir do estudo?

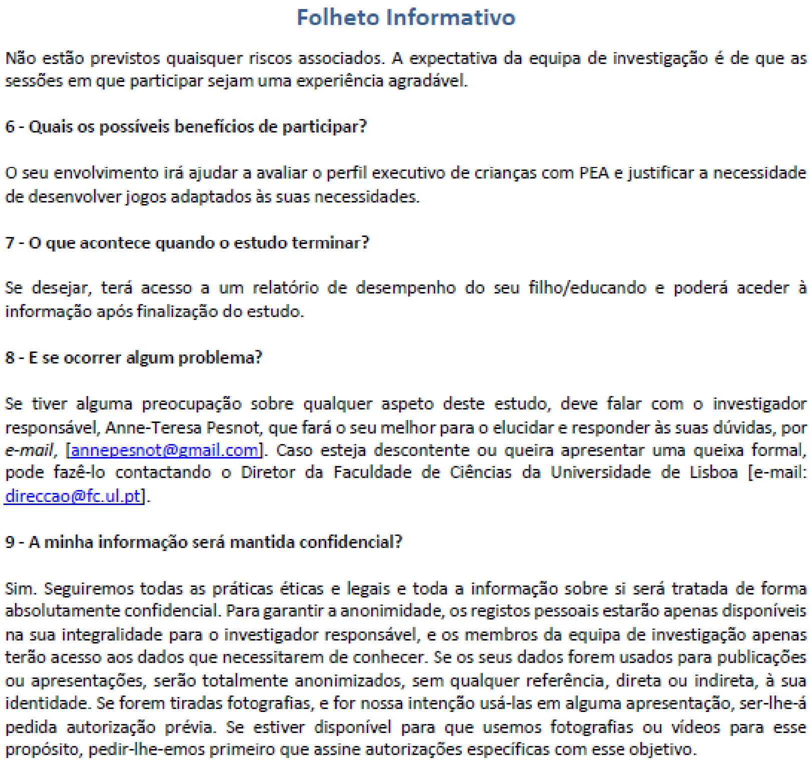
É livre de desistir, em qualquer altura, sem ter que fornecer quaisquer razões ou explicações.

4 - O que terei de fazer no âmbito do estudo?

No caso em que se proceda ao teste da aplicação computacional desenvolvida serão recolhidos dados acerca do olhar da criança e ao seu desempenho no jogo, fazendo recurso a um sensor biométrico.

5 - Quais as desvantagens e riscos de participar?

Figure 3.17: Informative leaflet (page 1).



Folheto Informativo

Não estão previstos quaisquer riscos associados. A expectativa da equipa de investigação é de que as sessões em que participar sejam uma experiência agradável.

6 - Quais os possíveis benefícios de participar?

O seu envolvimento irá ajudar a avaliar o perfil executivo de crianças com PEA e justificar a necessidade de desenvolver jogos adaptados às suas necessidades.

7 - O que acontece quando o estudo terminar?

Se desejar, terá acesso a um relatório de desempenho do seu filho/educando e poderá aceder à informação após finalização do estudo.

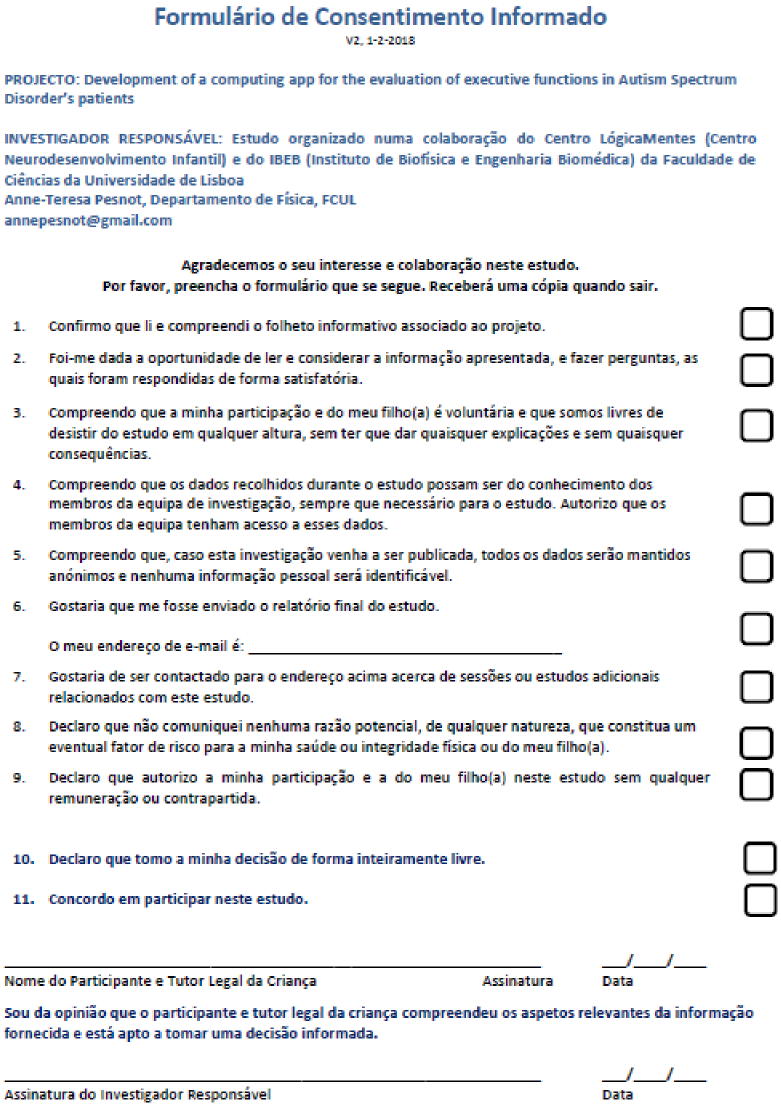
8 - E se ocorrer algum problema?

Se tiver alguma preocupação sobre qualquer aspeto deste estudo, deve falar com o investigador responsável, Anne-Teresa Pesnot, que fará o seu melhor para o elucidar e responder às suas dúvidas, por *e-mail*, [annepesnot@gmail.com]. Caso esteja descontente ou queira apresentar uma queixa formal, pode fazê-lo contactando o Diretor da Faculdade de Ciências da Universidade de Lisboa [e-mail: direccao@fc.ul.pt].

9 - A minha informação será mantida confidencial?

Sim. Seguiremos todas as práticas éticas e legais e toda a informação sobre si será tratada de forma absolutamente confidencial. Para garantir a anonimidade, os registos pessoais estarão apenas disponíveis na sua integralidade para o investigador responsável, e os membros da equipa de investigação apenas terão acesso aos dados que necessitarem de conhecer. Se os seus dados forem usados para publicações ou apresentações, serão totalmente anonimizados, sem qualquer referência, direta ou indireta, à sua identidade. Se forem tiradas fotografias, e for nossa intenção usá-las em alguma apresentação, ser-lhe-á pedida autorização prévia. Se estiver disponível para que usemos fotografias ou vídeos para esse propósito, pedir-lhe-emos primeiro que assine autorizações específicas com esse objetivo.

Figure 3.18: Informative leaflet (page 2).



Formulário de Consentimento Informado

V2, 1-2-2018

PROJECTO: Development of a computing app for the evaluation of executive functions in Autism Spectrum Disorder's patients

INVESTIGADOR RESPONSÁVEL: Estudo organizado numa colaboração do Centro LógicaMentes (Centro Neurodesenvolvimento Infantil) e do IBEB (Instituto de Biofísica e Engenharia Biomédica) da Faculdade de Ciências da Universidade de Lisboa
Anne-Teresa Pesnot, Departamento de Física, FCUL
annepesnot@gmail.com

Agradecemos o seu interesse e colaboração neste estudo.
Por favor, preencha o formulário que se segue. Receberá uma cópia quando sair.

1. Confirmo que li e compreendi o folheto informativo associado ao projeto. ☐
2. Foi-me dada a oportunidade de ler e considerar a informação apresentada, e fazer perguntas, as quais foram respondidas de forma satisfatória. ☐
3. Compreendo que a minha participação e do meu filho(a) é voluntária e que somos livres de desistir do estudo em qualquer altura, sem ter que dar quaisquer explicações e sem quaisquer consequências. ☐
4. Compreendo que os dados recolhidos durante o estudo possam ser do conhecimento dos membros da equipa de investigação, sempre que necessário para o estudo. Autorizo que os membros da equipa tenham acesso a esses dados. ☐
5. Compreendo que, caso esta investigação venha a ser publicada, todos os dados serão mantidos anónimos e nenhuma informação pessoal será identificável. ☐
6. Gostaria que me fosse enviado o relatório final do estudo. ☐

   O meu endereço de e-mail é: ______________________________
7. Gostaria de ser contactado para o endereço acima acerca de sessões ou estudos adicionais relacionados com este estudo. ☐
8. Declaro que não comuniquei nenhuma razão potencial, de qualquer natureza, que constitua um eventual fator de risco para a minha saúde ou integridade física ou do meu filho(a). ☐
9. Declaro que autorizo a minha participação e a do meu filho(a) neste estudo sem qualquer remuneração ou contrapartida. ☐
10. **Declaro que tomo a minha decisão de forma inteiramente livre.** ☐
11. **Concordo em participar neste estudo.** ☐

____ ___/___/___
Nome do Participante e Tutor Legal da Criança Assinatura Data

Sou da opinião que o participante e tutor legal da criança compreendeu os aspetos relevantes da informação fornecida e está apto a tomar uma decisão informada.

____ ___/___/___
Assinatura do Investigador Responsável Data

Figure 3.19: Written informed consent.

D) CERTIFICATE OF PARTICIPATION

CERTIFICADO DE PARTICIPAÇÃO

PARABÉNS

Agradecimento pela participação no estudo
“Development of a computing app for the evaluation of executive functions in Autism Spectrum Disorder’s patients”
Atribuído a __ de julho 2019

DRA. CLÁUDIA BANDEIRA DE LIMA

PROF. HUGO FERREIRA

ANNE-TERESA PESNOT

Figure 3.20: Certificate of participation offered to each participant.

E) HEATMAPS

Typical Development

Subject 1

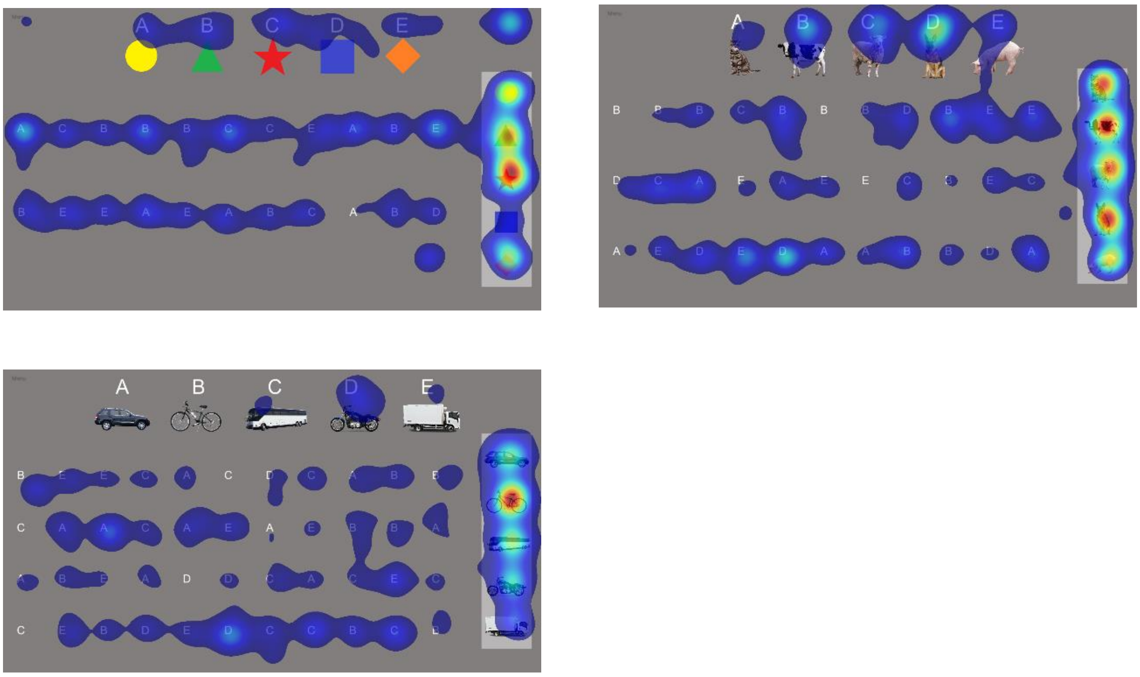


Figure 4.13: Heatmaps for the three levels of the Coding task.

Figure 4.14: Heatmaps for the three levels of the Puzzles task.

Subject 2

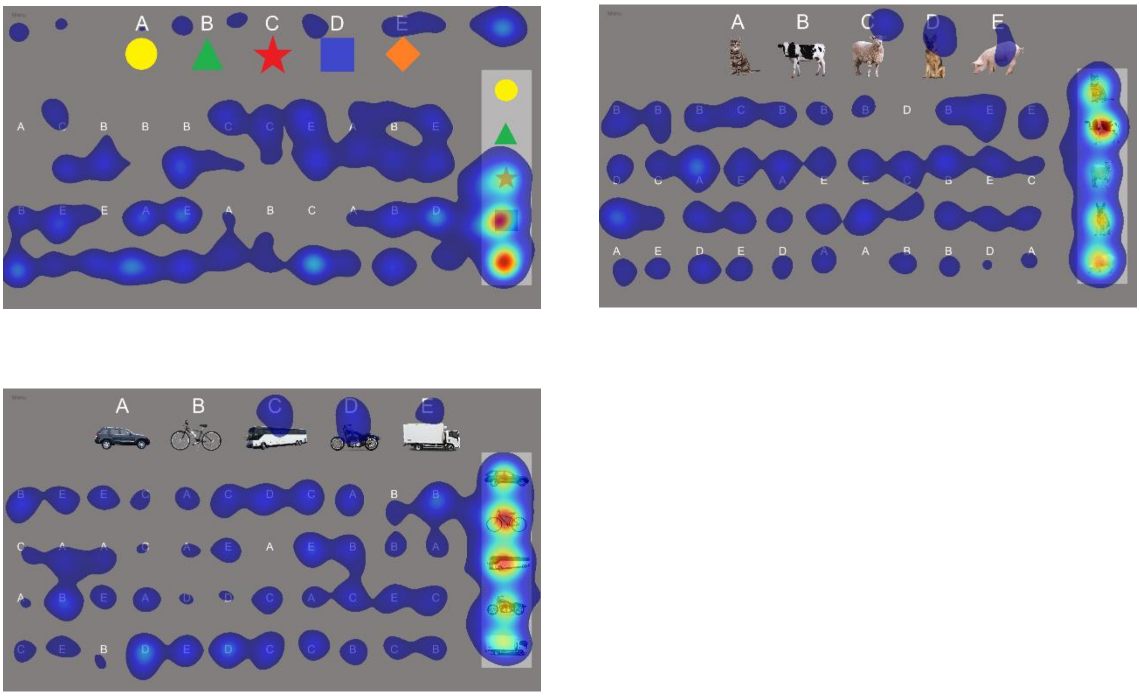


Figure 4.14: Heatmaps for the three levels of the Coding task.

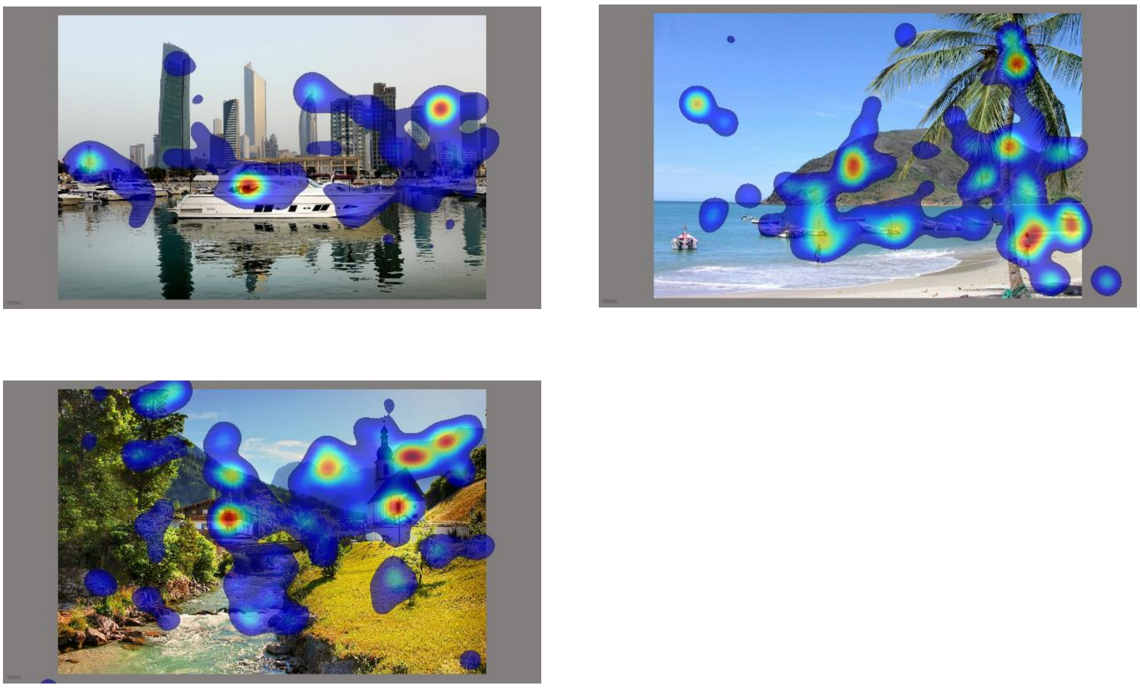


Figure 4.15: Heatmaps for the three levels of the Puzzles task.

Subject 3

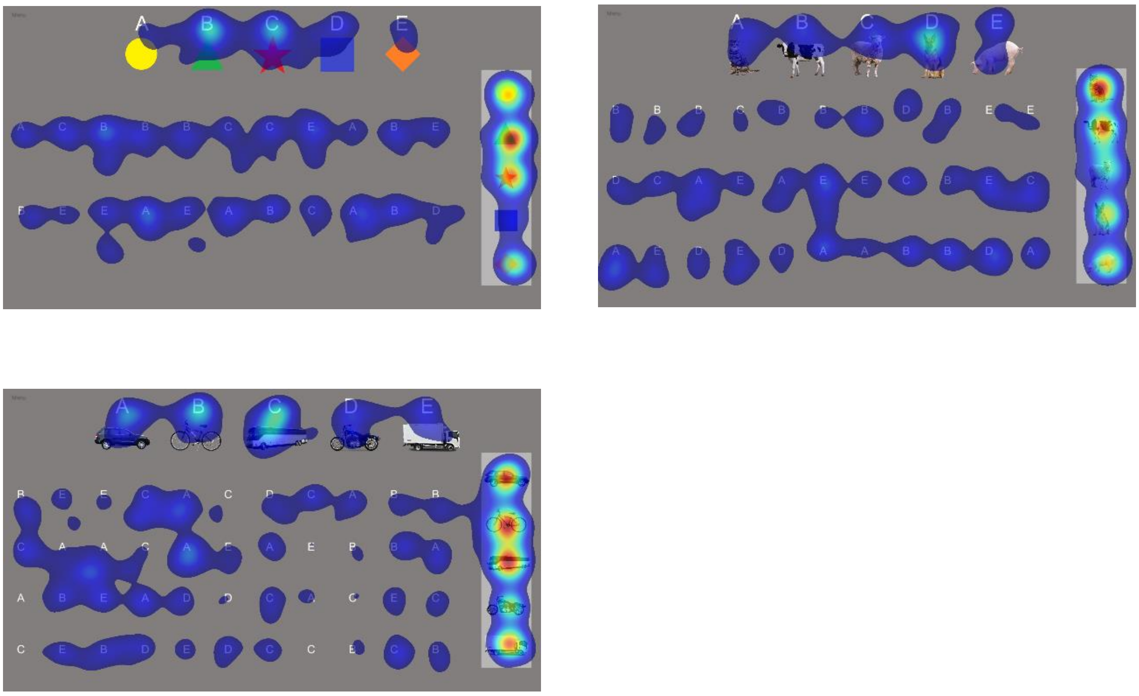


Figure 4.16: Heatmaps for the three levels of the Coding task.

4.17: Heatmaps for the three levels of the Puzzles task.

Subject 4

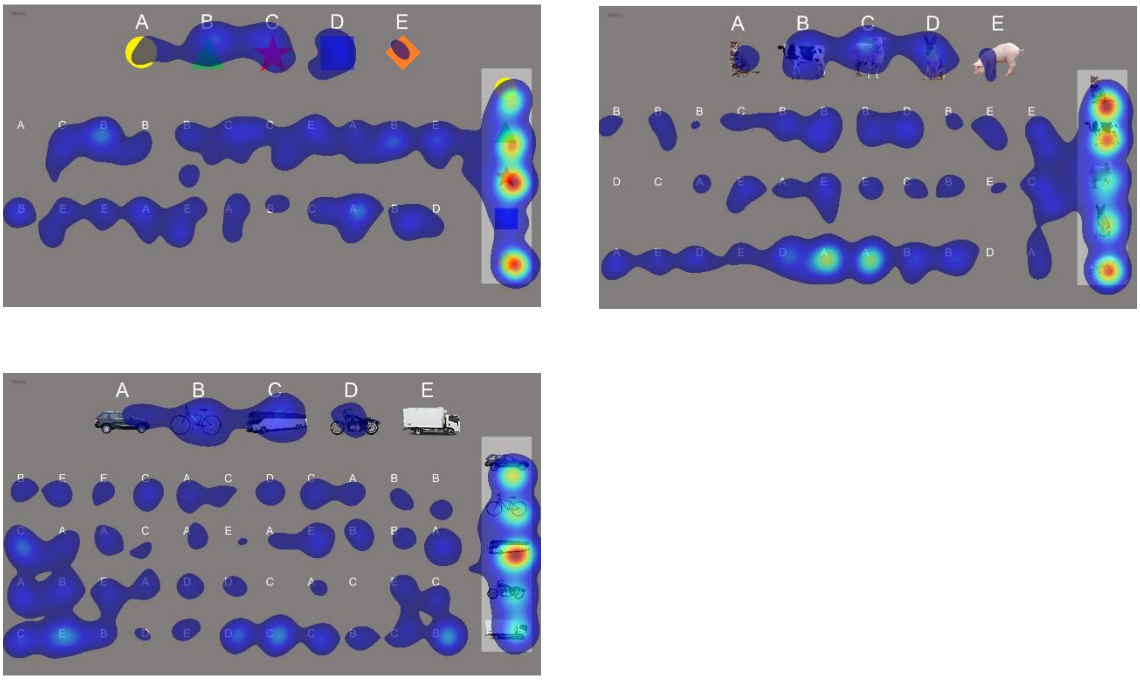


Figure 4.18: Heatmaps for the three levels of the Coding task.

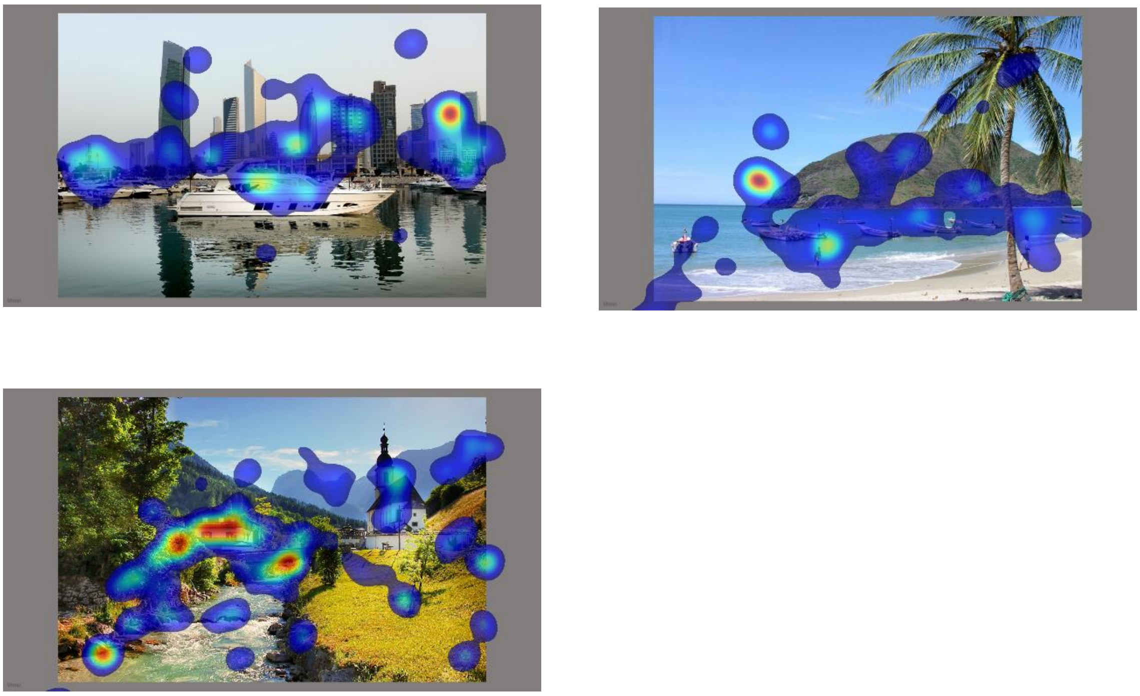


Figure 4.19: Heatmaps for the three levels of the Puzzles task.

Subject 5

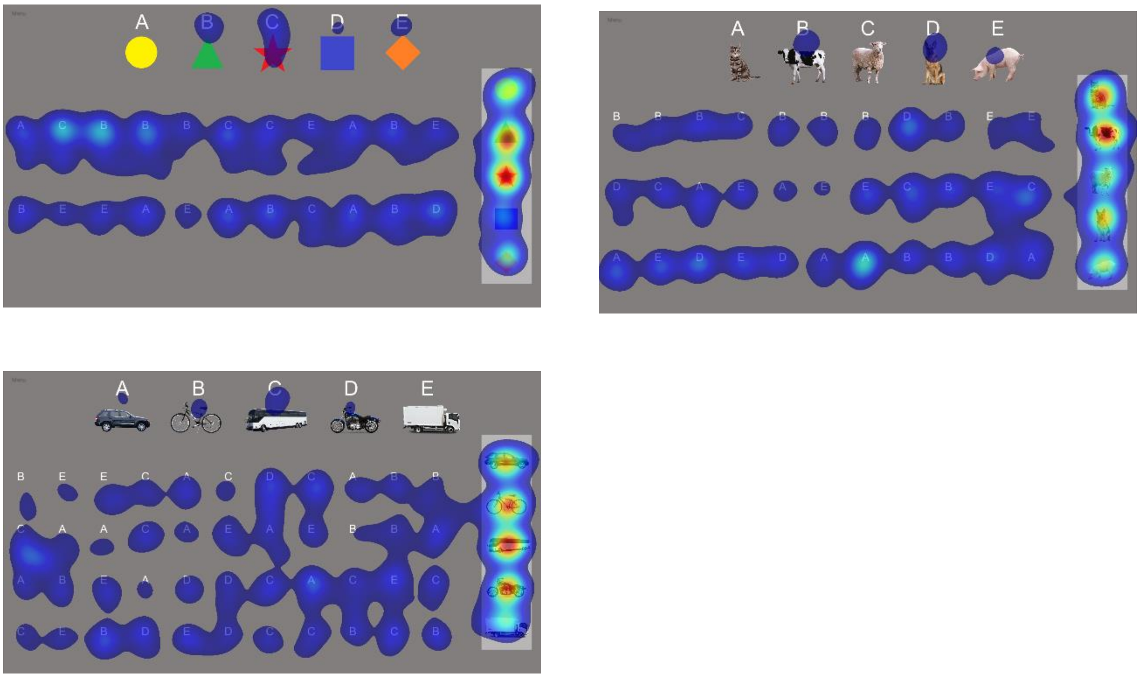


Figure 4.20: Heatmaps for the three levels of the Coding task.

Figure 4.21: Heatmaps for the three levels of the Puzzles task.

Subject 6

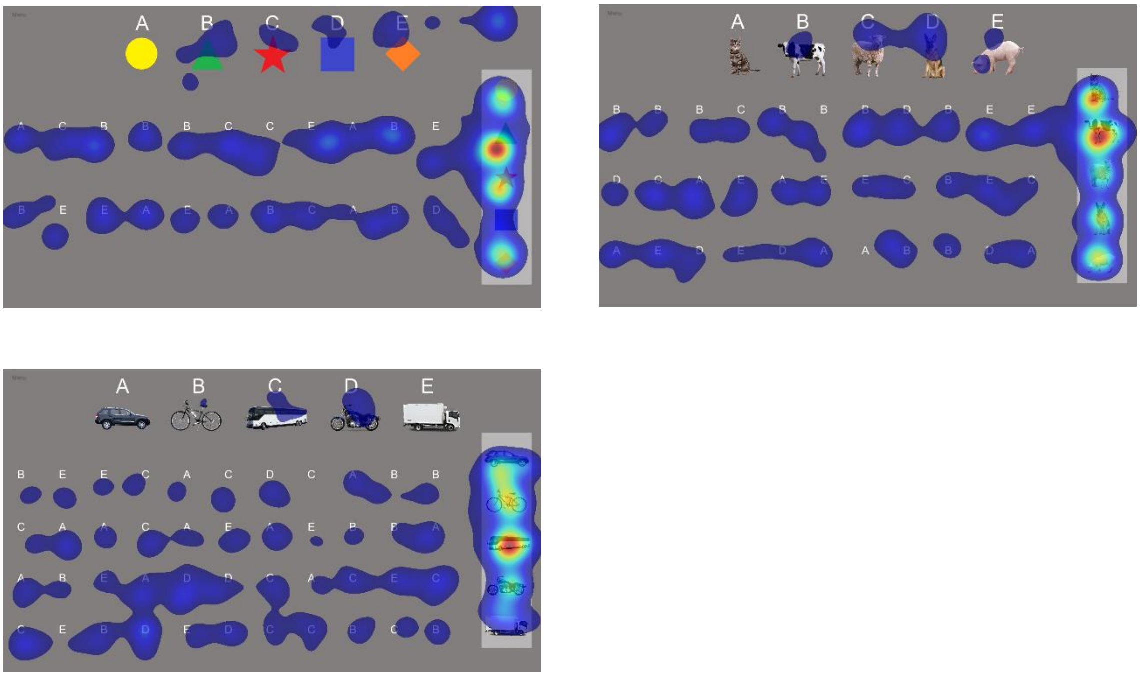


Figure 4.22: Heatmaps for the three levels of the Coding task.

Figure 4.23: Heatmaps for the three levels of the Puzzles task.

Subject 7

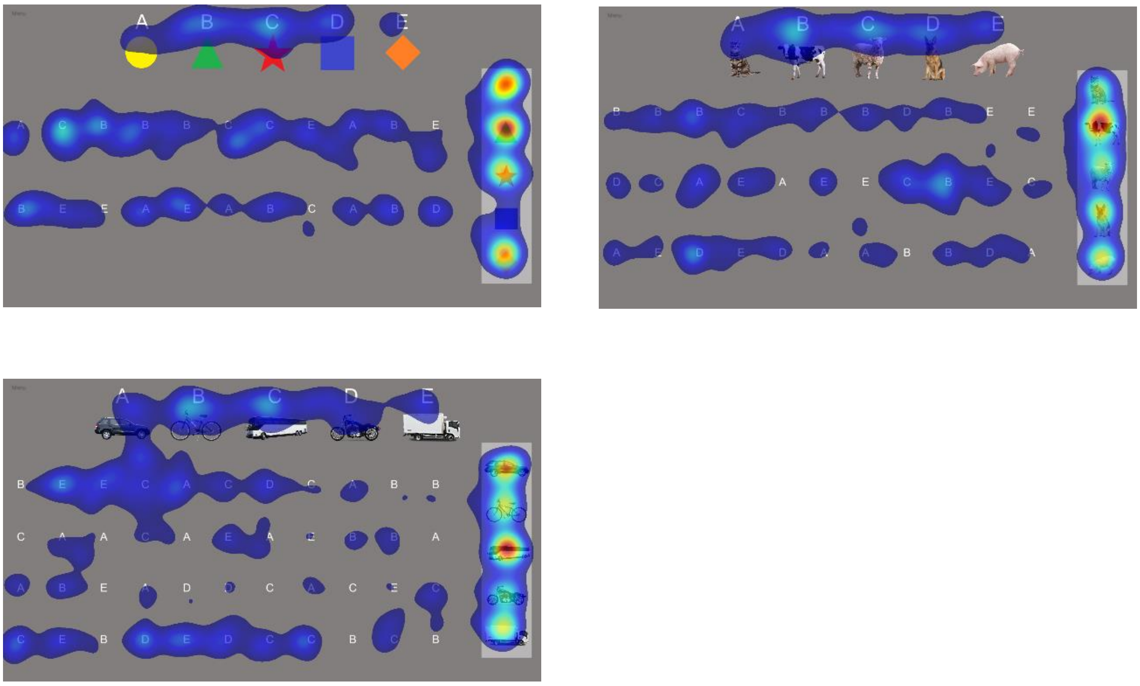


Figure 4.24: Heatmaps for the three levels of the Coding task.

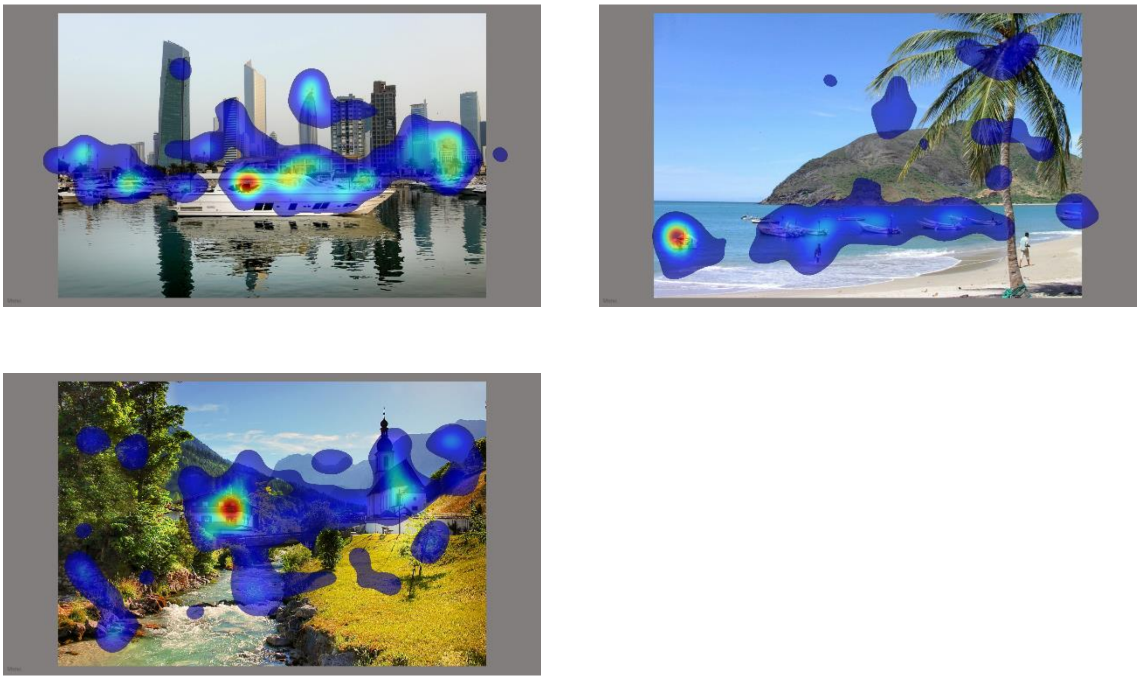


Figure 4.25: Heatmaps for the three levels of the Puzzles task.

Subject 8

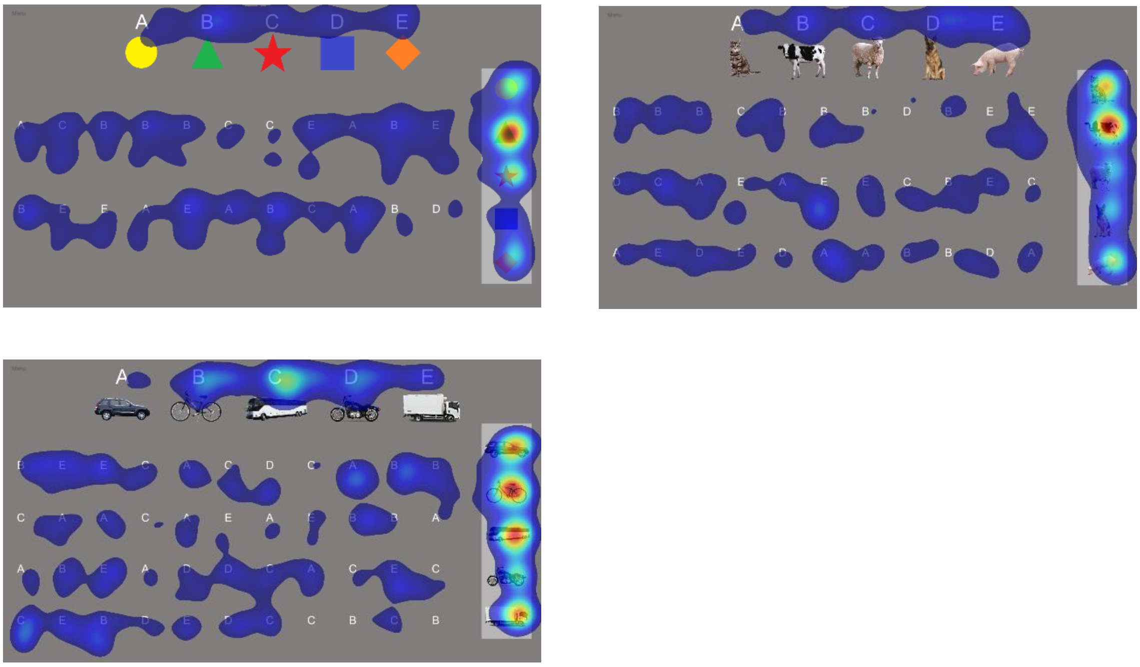


Figure 4.26: Heatmaps for the three levels of the Coding task.

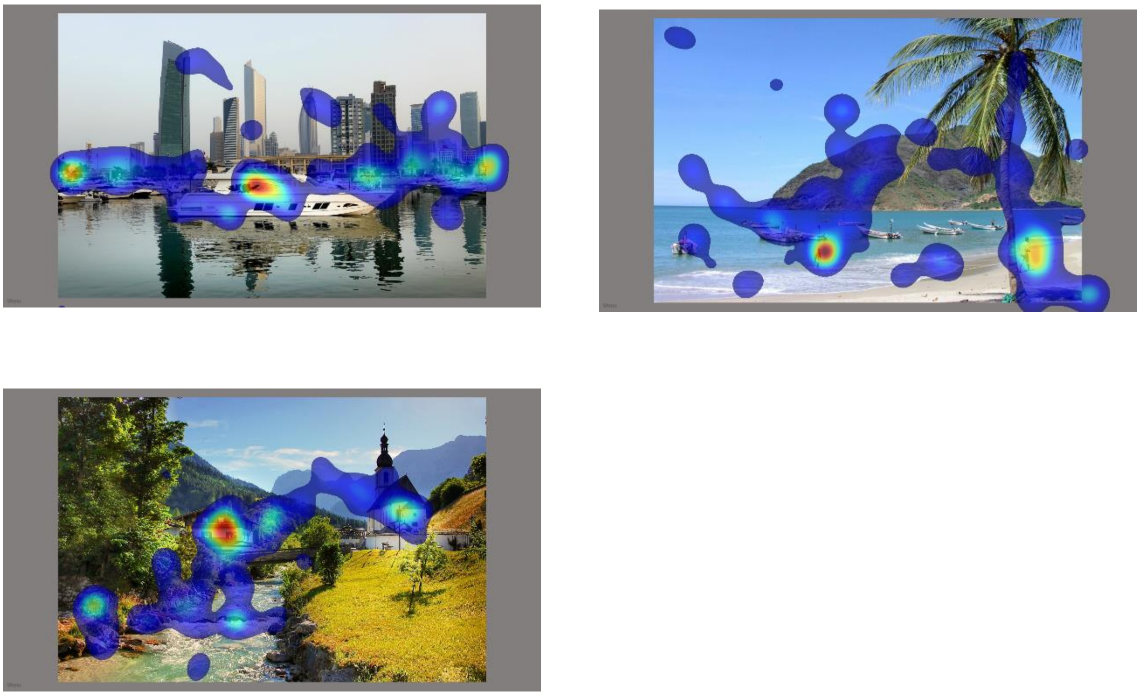


Figure 4.27: Heatmaps for the three levels of the Puzzles task.

Subject 9

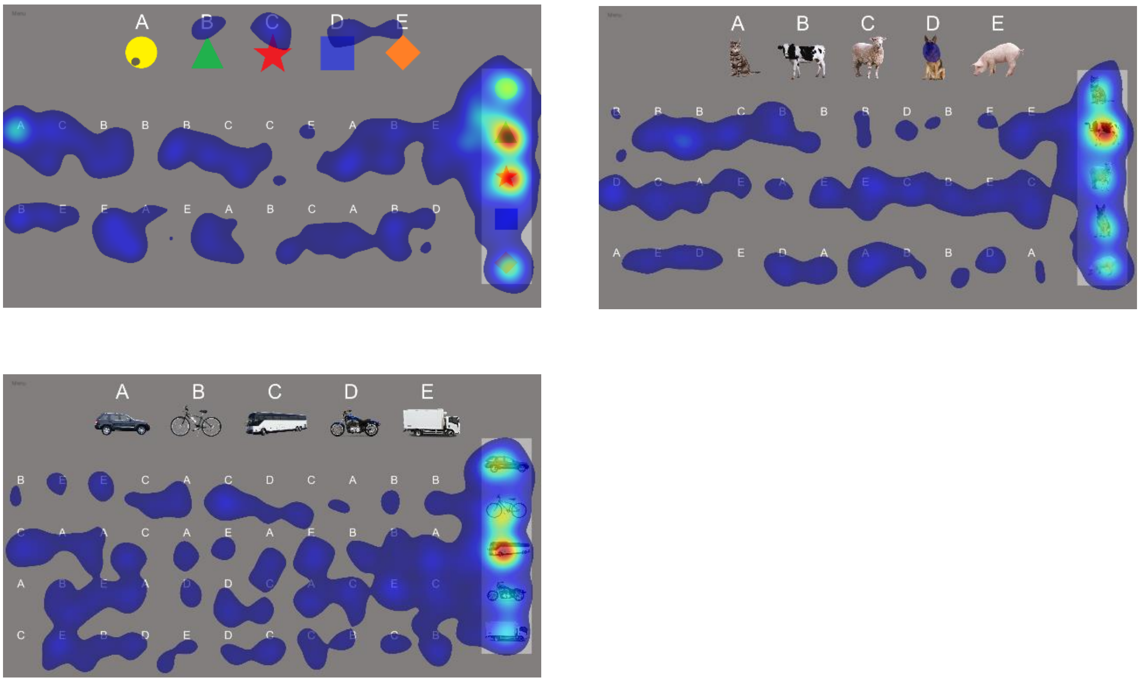


Figure 4.28: Heatmaps for the three levels of the Coding task.

Figure 4.29: Heatmaps for the three levels of the Coding task.

Subject 11

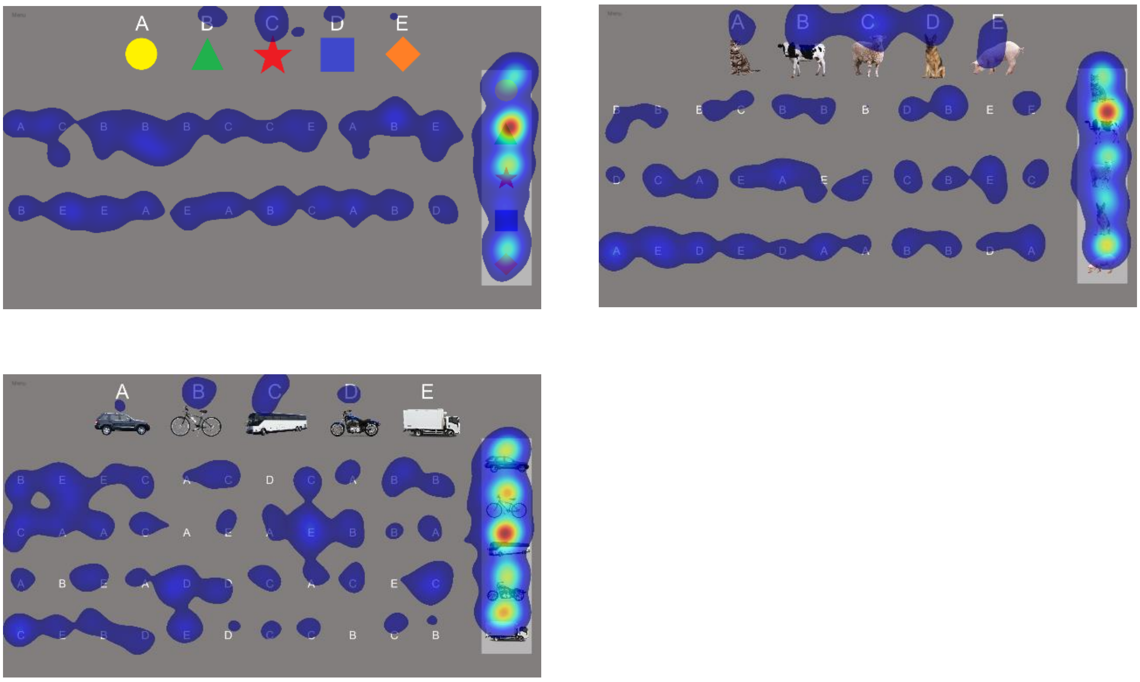


Figure 4.30: Heatmaps for the three levels of the Coding task.

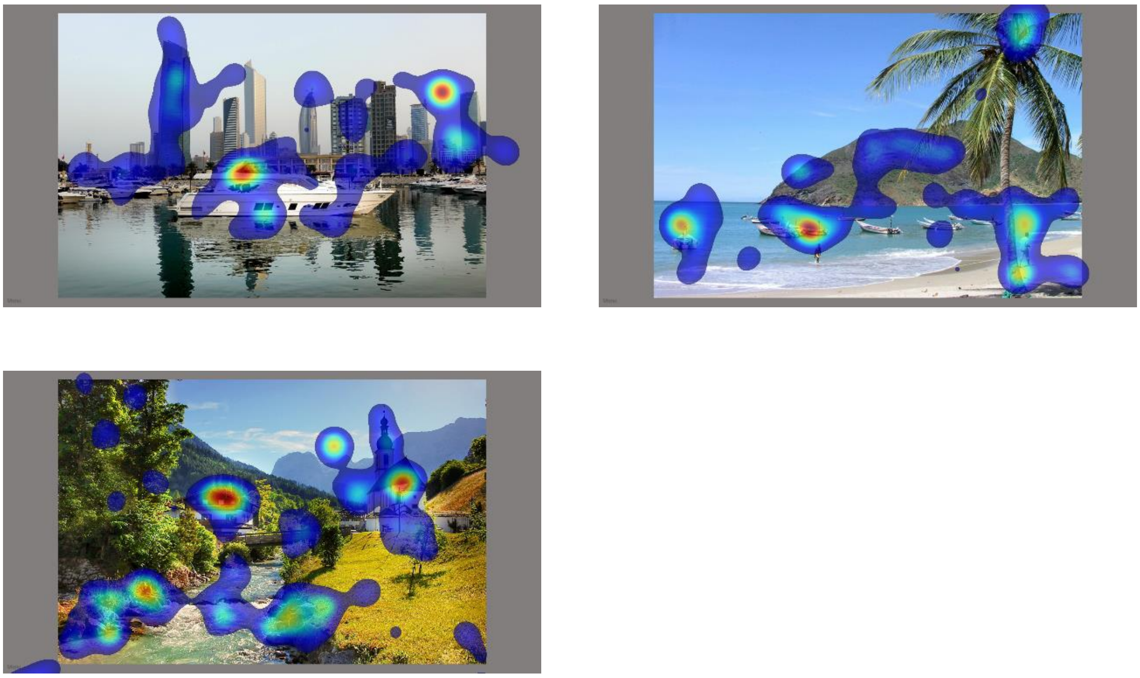


Figure 4.31: Heatmaps for the three levels of the Puzzles task.

Subject 12

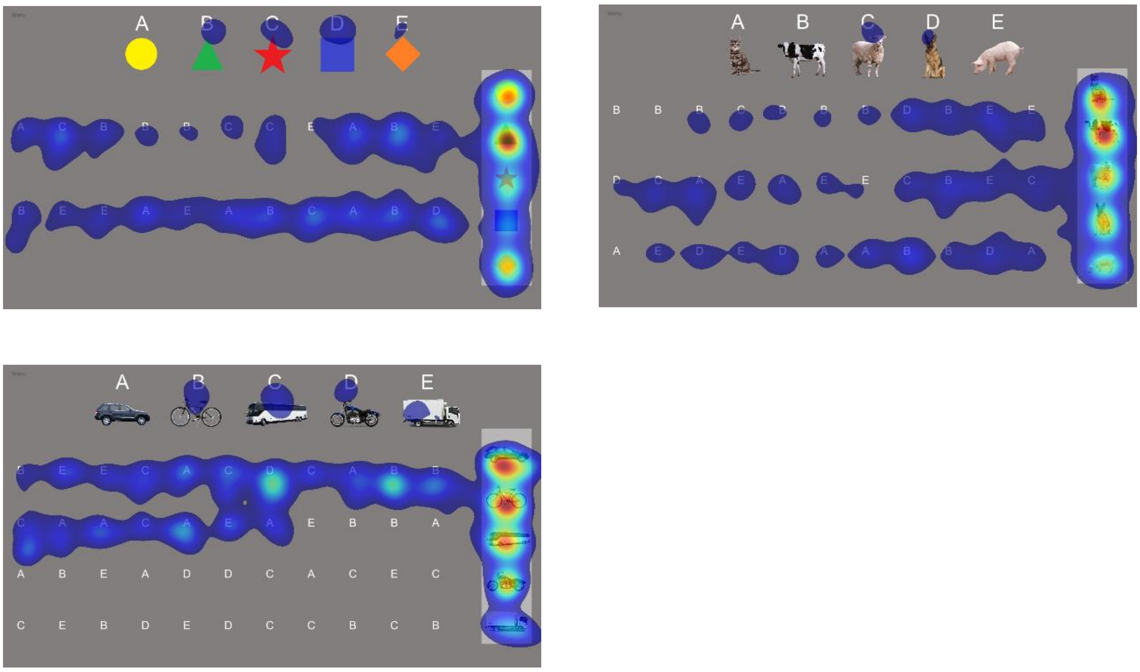


Figure 4.32: Heatmaps for the three levels of the Coding task.

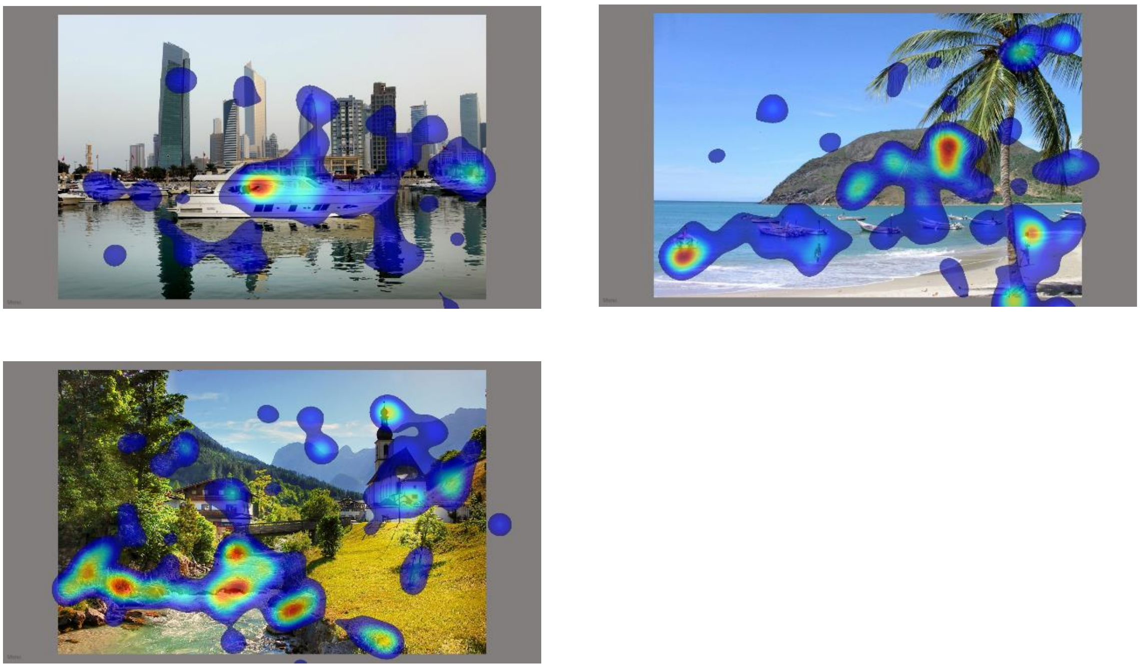


Figure 4.33: Heatmaps for the three levels of the Puzzles task.

Subject 13

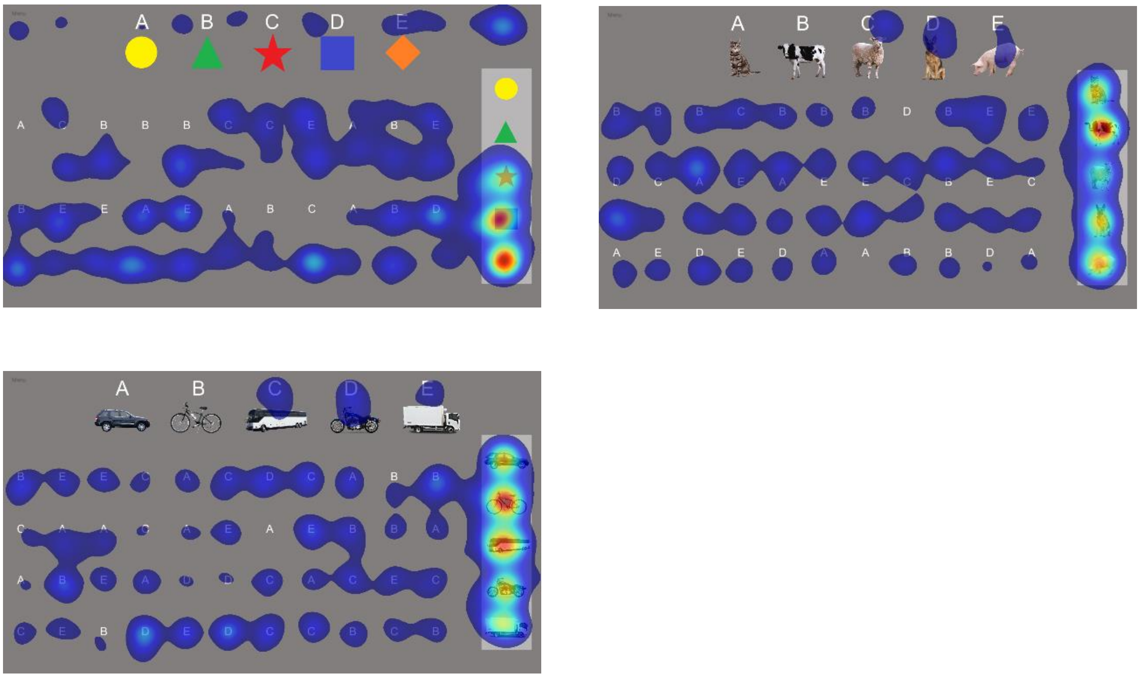


Figure 4.34: Heatmaps for the three levels of the Coding task.

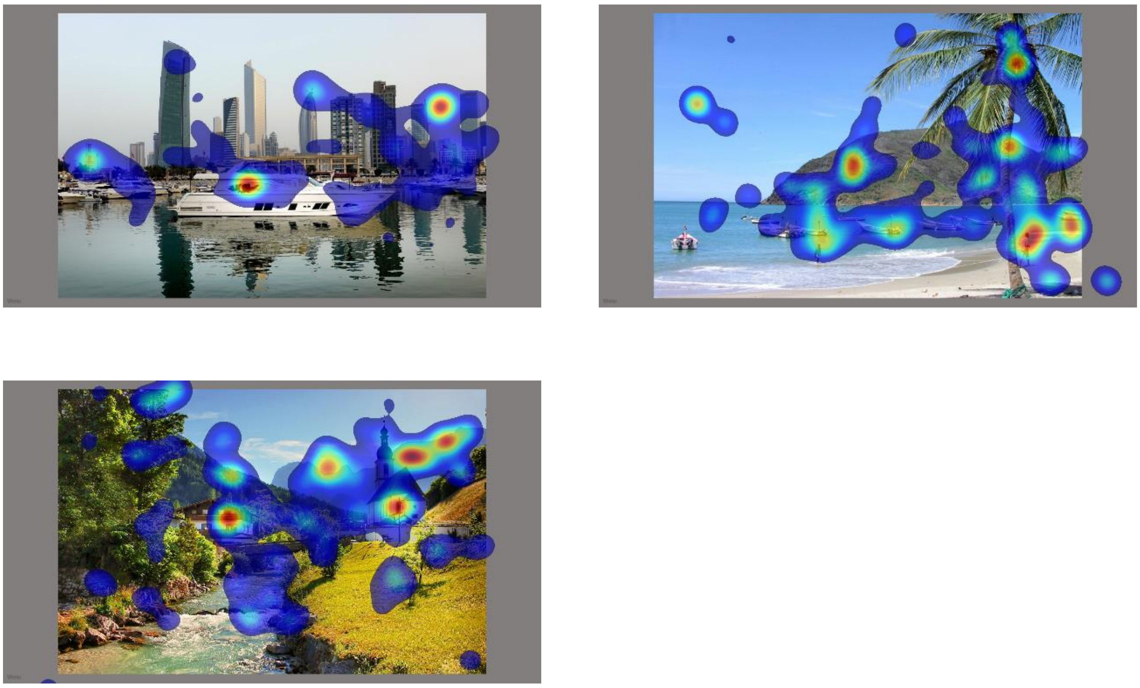


Figure 4.35: Heatmaps for the three levels of the Puzzles task.

Subject 14

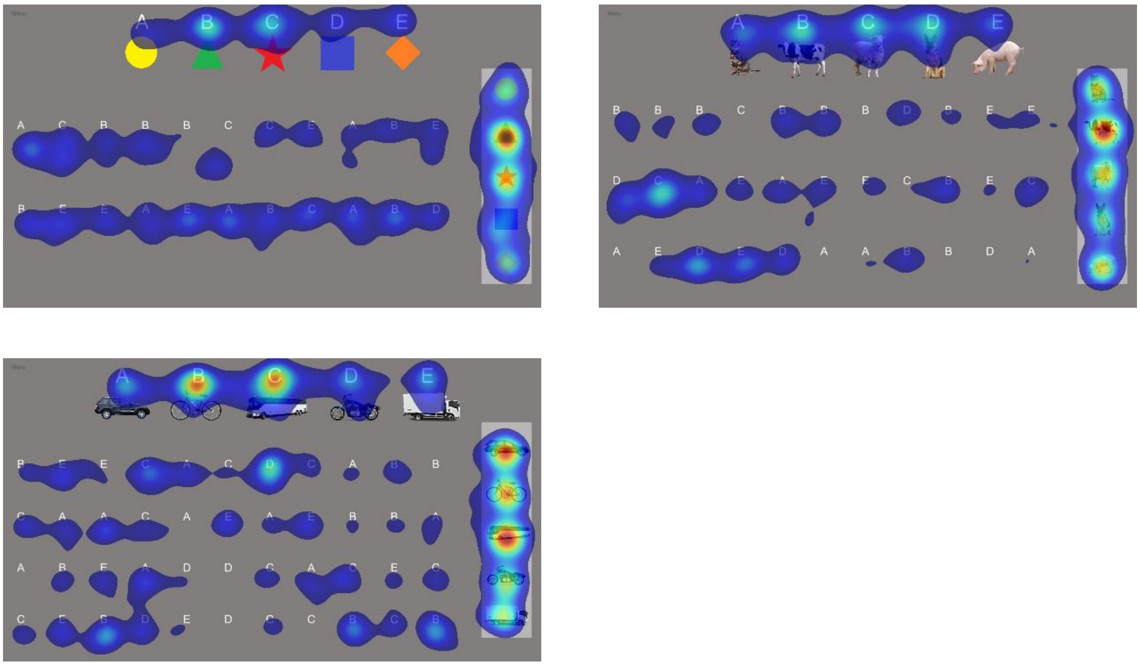


Figure 4.36: Heatmaps for the three levels of the Coding task.

Figure 4.37: Heatmaps for the three levels of the Puzzles task.

ASD

Subject 1

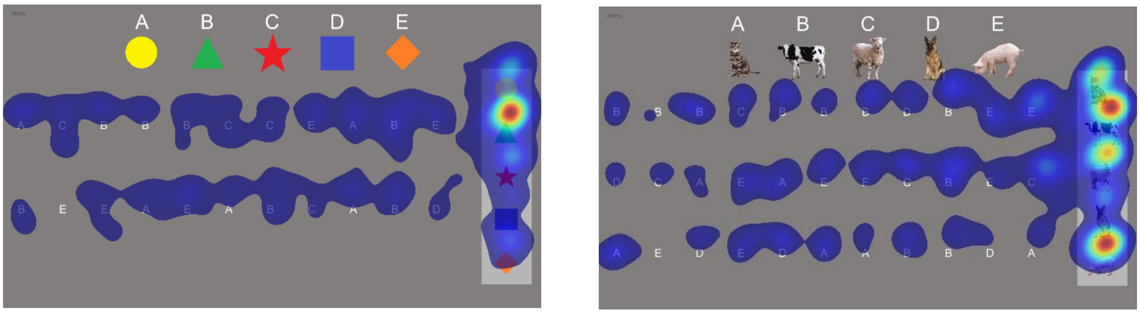


Figure 38: Heatmaps for levels 1 and 2 of the Coding task.

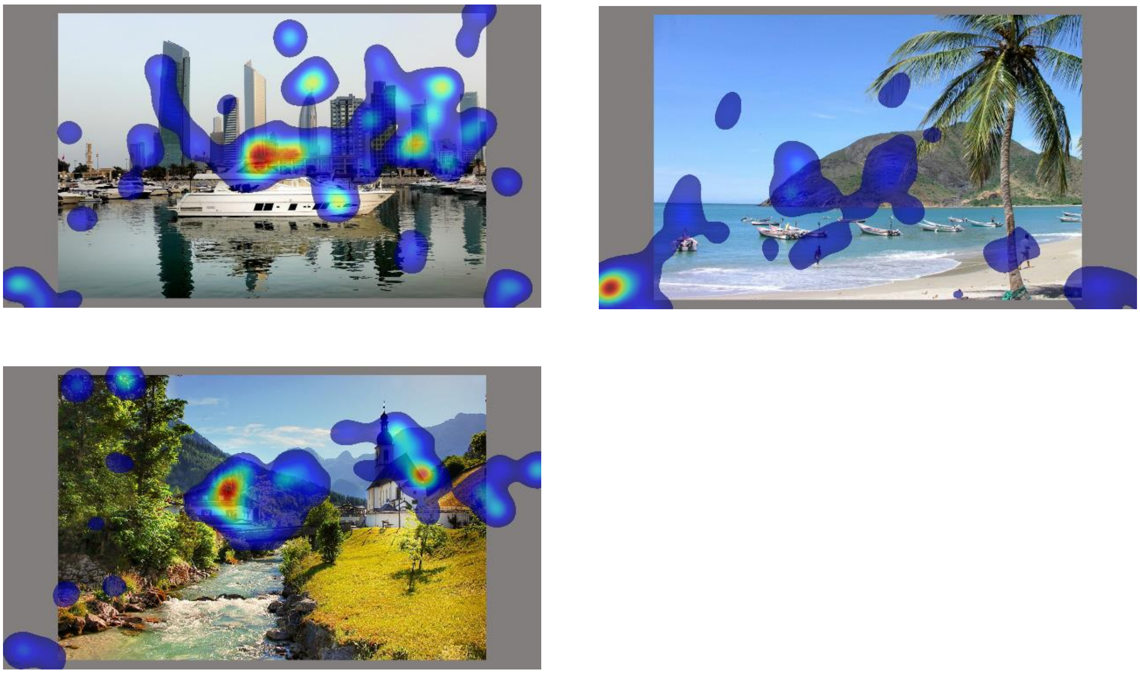


Figure 4.39: Heatmaps for the three levels of the Puzzles task.

Subject 2

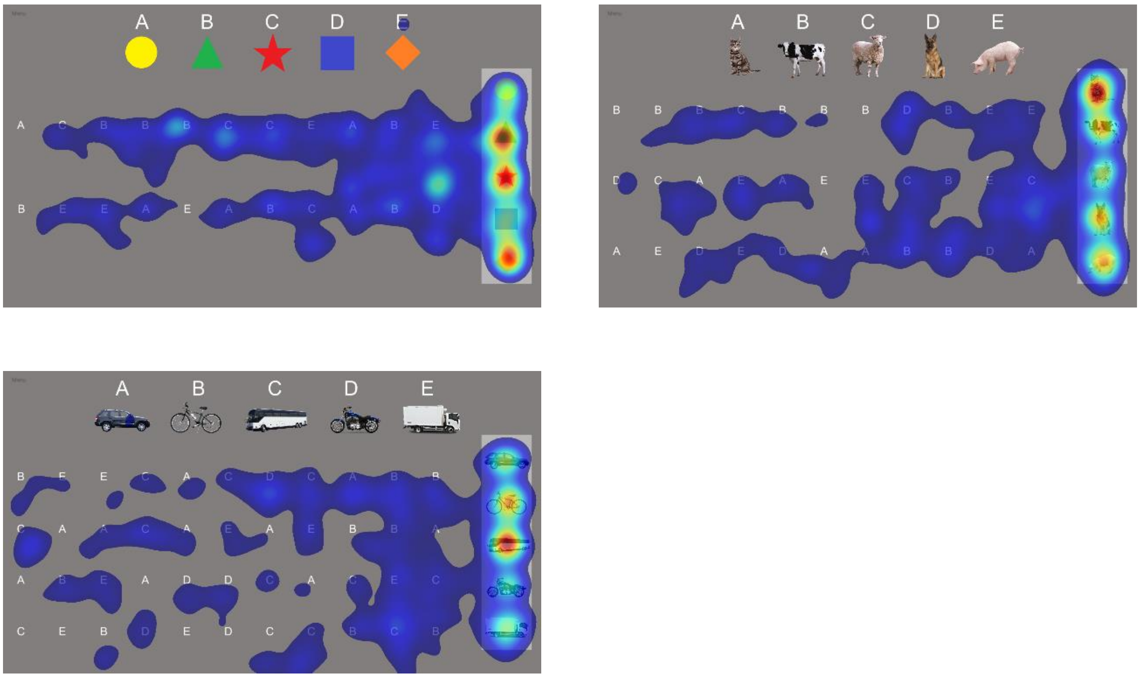


Figure 4.40: Heatmaps for the three levels of the Coding task.

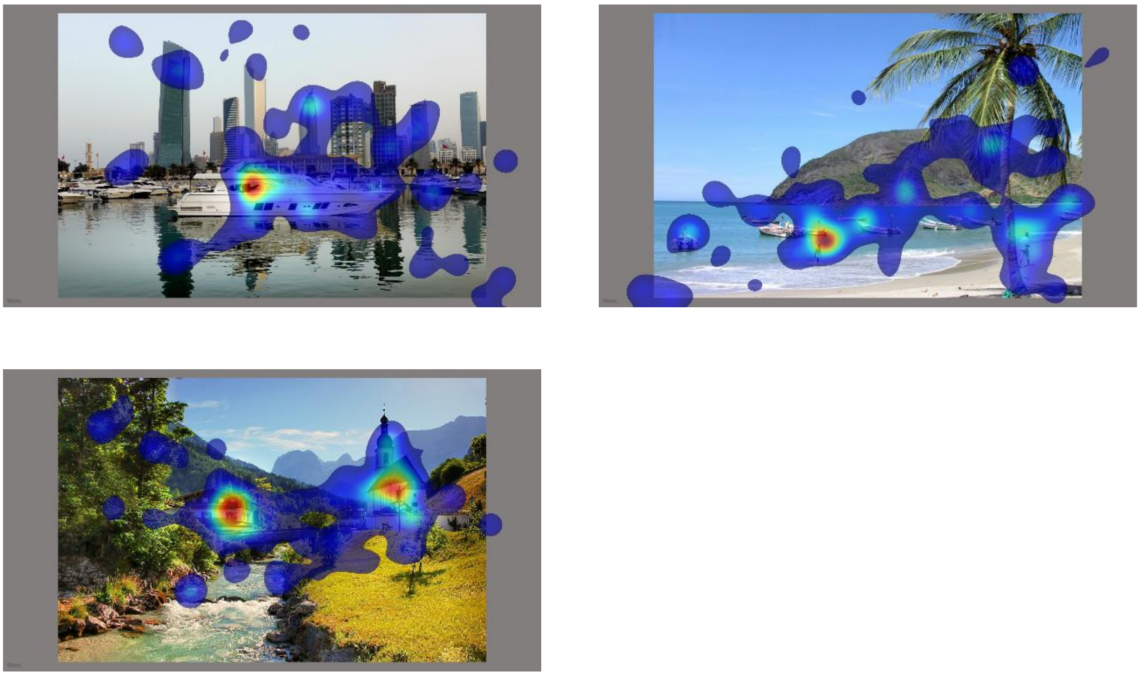


Figure 4.41: Heatmaps for the three levels of the Puzzles task.

Subject 3

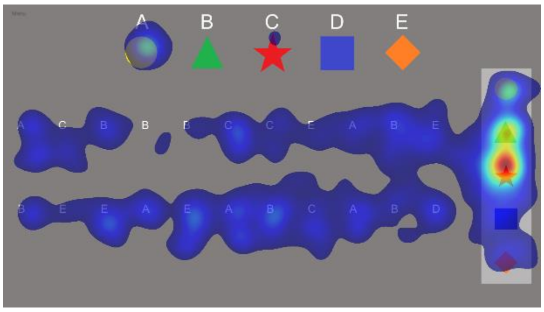


Figure 4.42: Heatmaps for level 1of the Coding task.

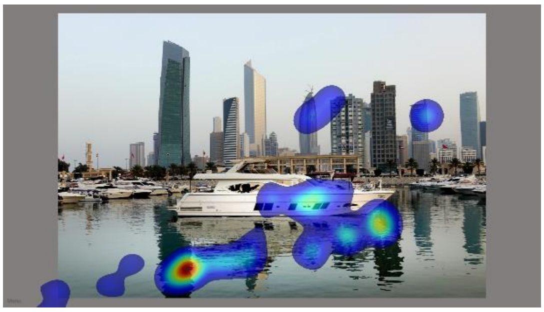


Figure 4.43: Heatmaps for level 1 of the Puzzles task.

Subject 4

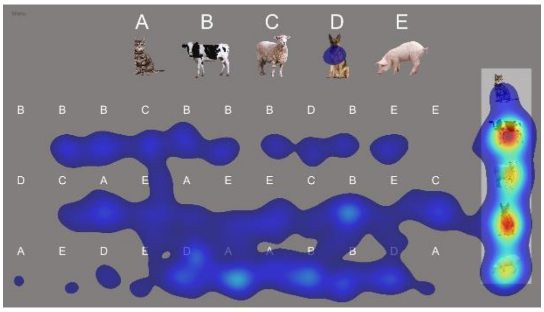


Figure 4.44: Heatmap for level 2 of the Coding task.

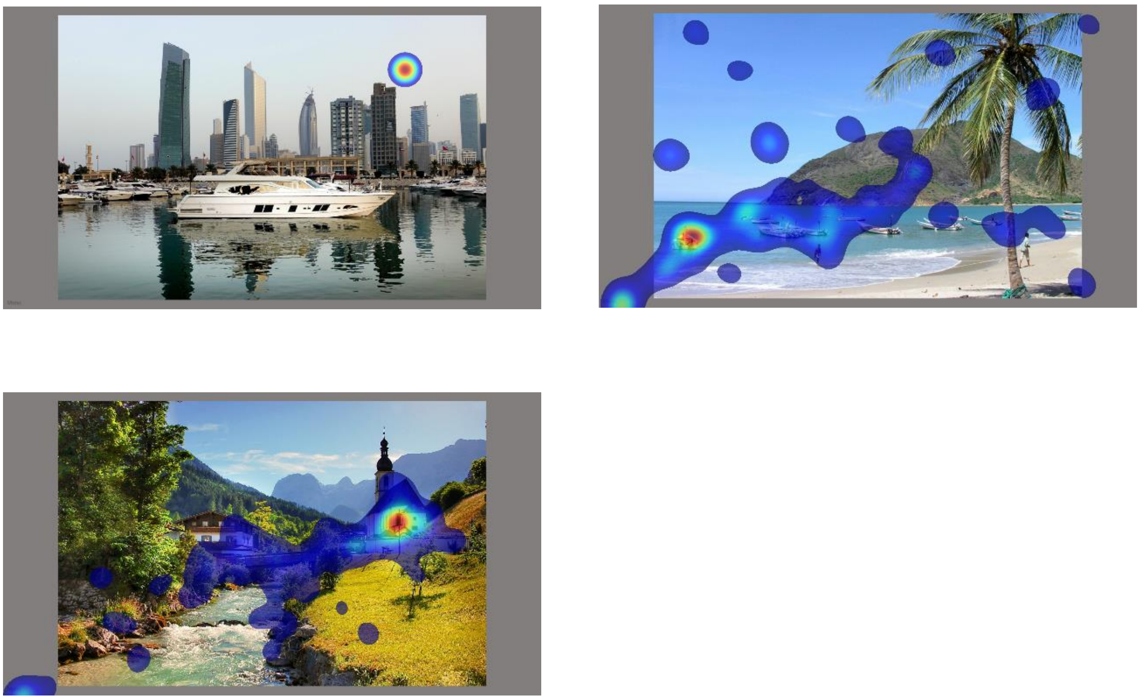


Figure 4.45: Heatmaps for the three levels of the Coding task.

Subject 5

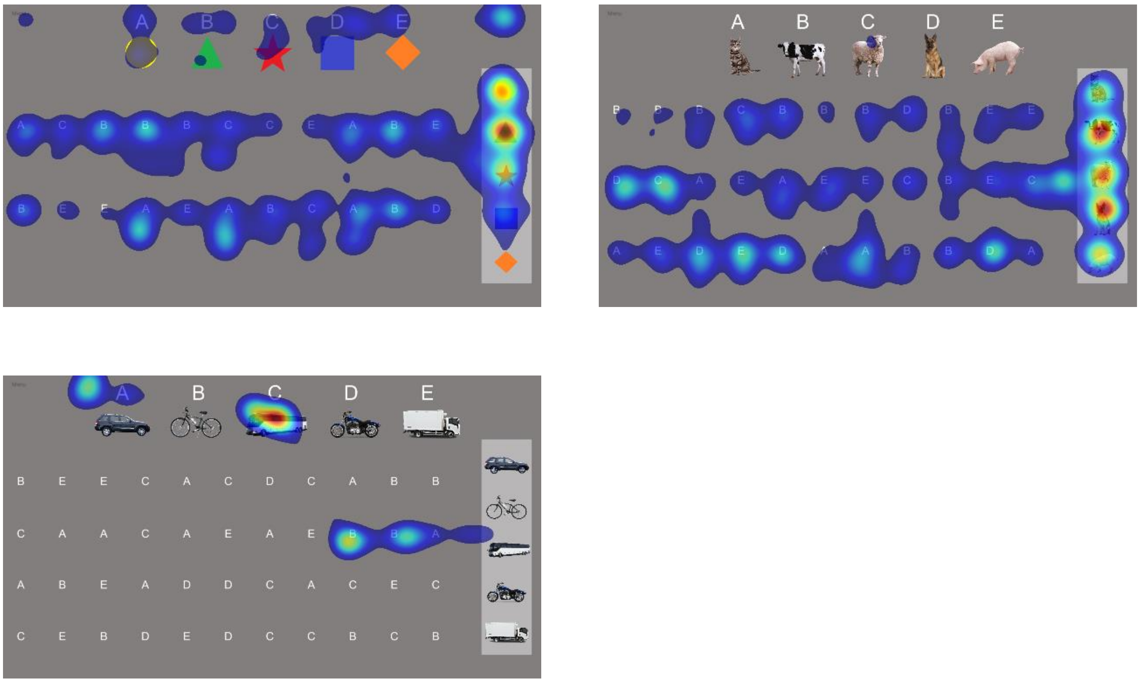


Figure 4.46: Heatmaps for the three levels of the Coding task.

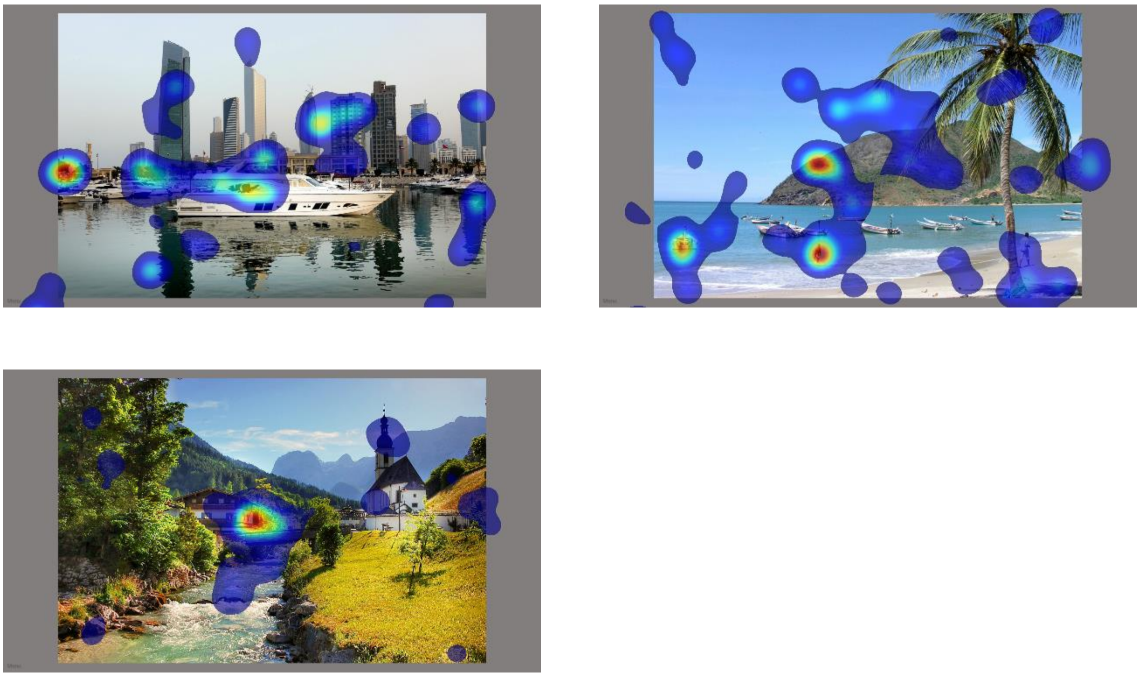


Figure 4.47: Heatmaps for the three levels of the Puzzles task.

Subject 6

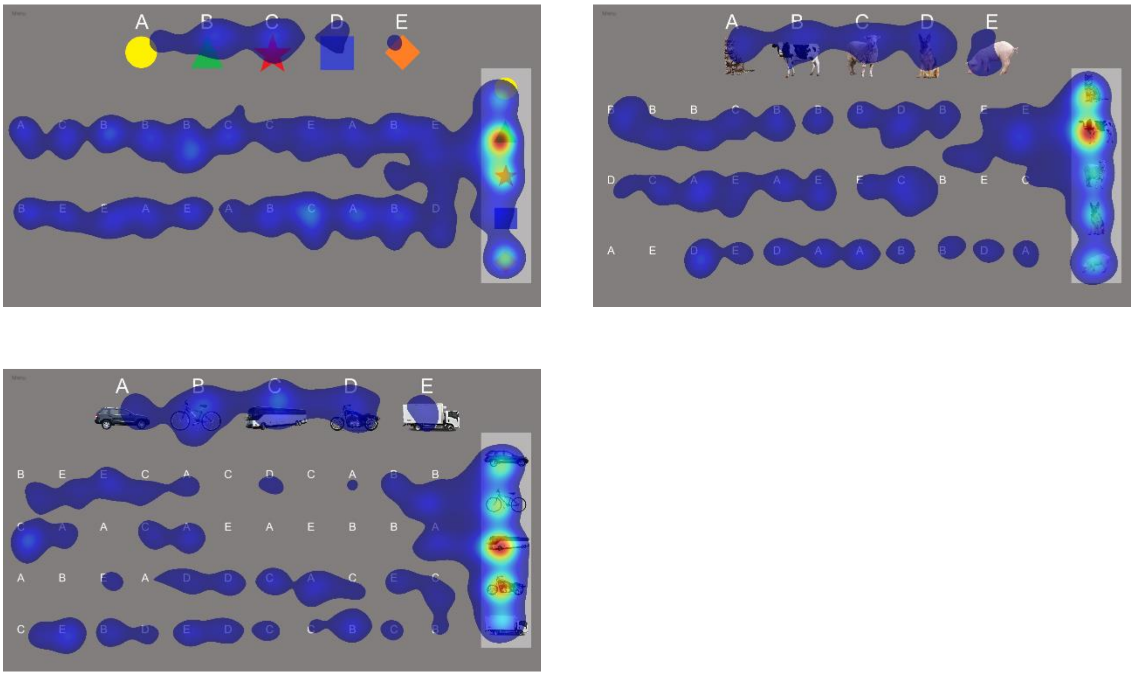


Figure 4.48: Heatmaps for the three levels of the Coding task.

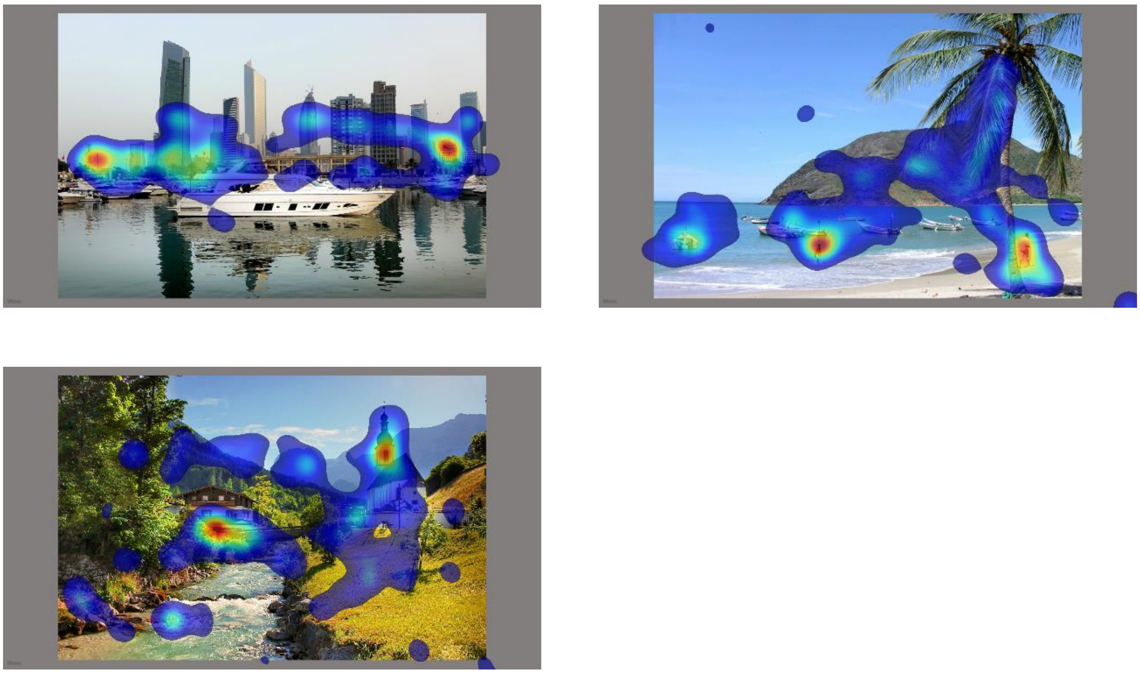


Figure 4.49: Heatmaps for the three levels of the Coding task.

Subject 8

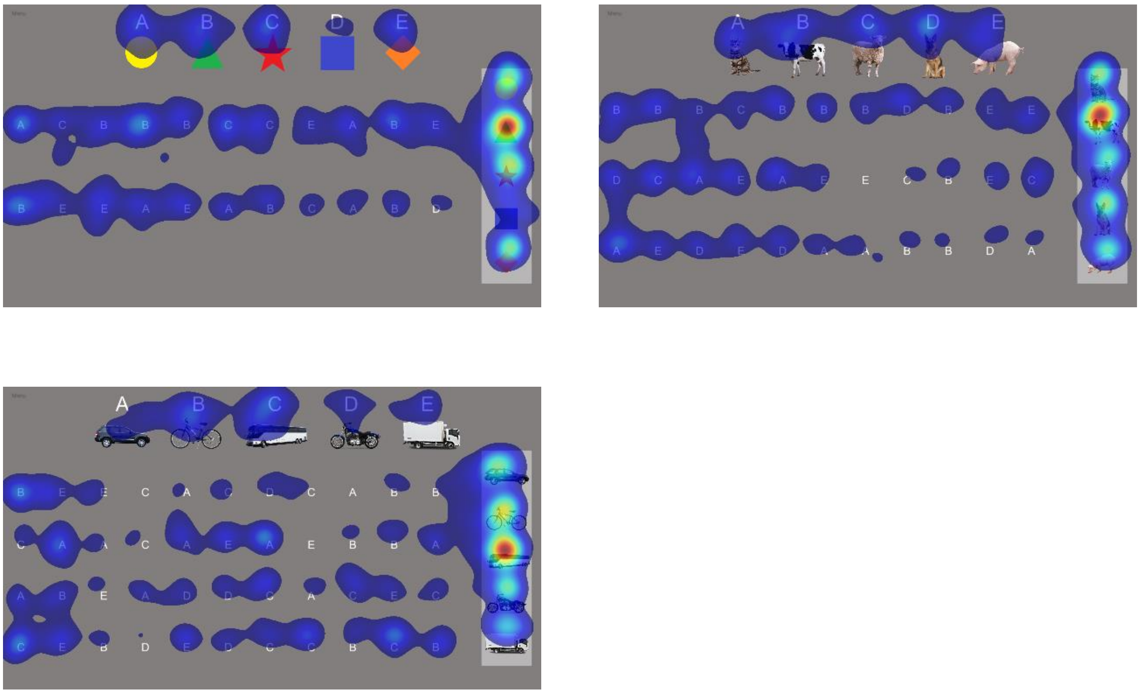


Figure 4.50: Heatmaps for the three levels of the Coding task.

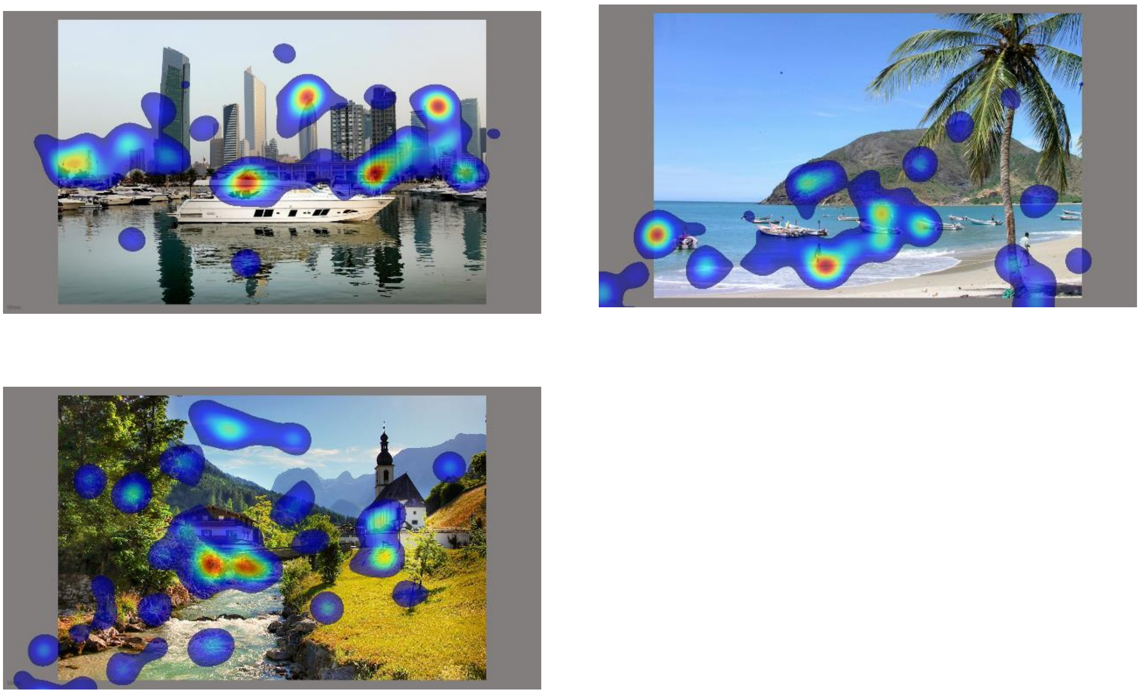


Figure 4.51: Heatmaps for the three levels of the Puzzles task.