COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

The Influence of Social Context on Sketching Behavior: A Study on Human- Robot Interaction

MASTER THESIS

2024 Hillary Pedrizzi

COMENIUS UNIVERSITY IN BRATISLAVA

FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

The Influence of Social Context on Sketching Behavior: A Study on Human- Robot Interaction

MASTER THESIS

Study Programme: Cognitive Science

Field of study: 2503 Cognitive Science

Department: Department of Applied Informatics

Supervisor: Mgr. Xenia Daniela Poslon, PhD.

Consultant: Carlo Mazzola, PhD.

Comenius University Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Hillary Pedrizzi
Study programme:	Cognitive Science (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: The Influence of Social Context on Sketching Behavior: A Study on Human-Robot Interaction

Annotation: Graphic illustration through sketches reveals the private world of mental representations, involving complex cognitive processes such as perception, memory, motor skills, and social mechanisms. Different drawing styles emerge when individuals sketch a specific exemplar versus a general category of an object, and these representations evolve when repeatedly depicted within the same social context.

Aim: This thesis aims to investigate how social context, specifically the presence of a child robot, influences sketching behavior. The study seeks to understand the impact of joint attention on human drawing activities during human-robot interaction. Objectives include defining the complexity in drawing, testing stimuli based on this complexity, and analyzing changes in sketching behavior in the presence of a robot observer.

Literature:
Fan, J.E. et al. (2023) 'Drawing as a versatile cognitive tool', Nature Reviews Psychology, 2(9), pp. 556–568.
Hawkins, R.D. et al. (2023) 'Visual resemblance and interaction history jointly constrain pictorial meaning', Nature Communications, 14(1).
Yang, J. and Fan, J. (2021) 'Visual communication of object concepts at different levels of abstraction', Journal of Vision, 21(9), p. 2951.

Supervisor:	Mgr. Xenia Daniela Poslon, PhD.
Consultant:	Dr. Carlo Mazzola
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	doc. RNDr. Tatiana Jajcayová, PhD.
Assigned:	25.05.2023

Approved: 25.05.2023 prof. Ing. Igor Farkaš, Dr.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Hillary Pedrizzi
Študijný program: kognitívna veda (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: informatika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: The Influence of Social Context on Sketching Behavior: A Study on Human-Robot Interaction
Vplyv sociálneho kontextu na kreslenie: Štúdia interakcie človeka s robotom

Anotácia: Grafická ilustrácia prostredníctvom náčrtov odhaľuje súkromný svet mentálnych reprezentácií, ktorý zahŕňa komplexné kognitívne procesy, ako je vnímanie, pamäť, motorické zručnosti a sociálne mechanizmy. Rozdielne štýly kreslenia sa objavujú, keď jednotlivci skicujú konkrétny exemplár v porovnaní so všeobecnou kategóriou objektu, a tieto reprezentácie sa vyvíjajú pri opakovanom zobrazovaní v rovnakom sociálnom kontexte.

Cieľ: Cieľom tejto práce je preskúmať, ako sociálny kontext, konkrétne prítomnosť detského robota, ovplyvňuje správanie pri kreslení. Cieľom práce bude skúmať vplyv spoločnej pozornosti na kreslenie človeka počas interakcie človek-robot. Medzi ciele patrí definovanie komplexnosti pri kreslení, testovanie podnetov založených na tejto komplexnosti a analýza zmien v správaní pri skicovaní v prítomnosti pozorovateľa robota.

Literatúra:
Fan, J.E. et al. (2023) 'Drawing as a versatile cognitive tool', Nature Reviews Psychology, 2(9), pp. 556–568.
Hawkins, R.D. et al. (2023) 'Visual resemblance and interaction history jointly constrain pictorial meaning', Nature Communications, 14(1).
Yang, J. and Fan, J. (2021) 'Visual communication of object concepts at different levels of abstraction', Journal of Vision, 21(9), p. 2951.

Vedúci: Mgr. Xenia Daniela Poslon, PhD.
Konzultant: Dr. Carlo Mazzola
Katedra: FMFI.KAI - Katedra aplikovanej informatiky
Vedúci katedry: doc. RNDr. Tatiana Jajcayová, PhD.

Dátum zadania: 25.05.2023

Dátum schválenia: 25.05.2023

prof. Ing. Igor Farkaš, Dr.
garant študijného programu

.....
študent

.....
vedúci práce

Declaration

I hereby declare that I have completed this master's thesis independently, without any unauthorized assistance. All literature and ideas sourced from other works have been properly cited.

Bratislava, 2024

Acknowledgments

I would like to express my gratitude to my supervisor, Mgr. Xenia D. Poslon, PhD, for imparting invaluable knowledge about various aspects of psychology. I am also deeply grateful to my consultant, Carlo Mazzola, PhD, for his insights into robotic bioengineering research. Their guidance has been instrumental throughout the entirety of this experiment. Additionally, I extend my sincere thanks to prof. Ing. Igor Farkaš, Dr. for facilitating my collaboration with the Italian Institute of Technology in Genoa, Italy. His support made my international research experience possible and significantly enriched my academic journey. The research leading to these results has received funding from the project titled TERAIS in the framework of the program Horizon-Widera-2021 of the European Union under the Grant agreement number 101079338.

Abstract

The work proposes investigating drawing activities in interactive contexts to shed light on the links between socio-cognitive and visual representation mechanisms. The creation of mental representations, a key area in cognitive science, can be explored through graphic illustrations like sketches, which are connected to human cognition, encompassing perception, memory, motor functions, and social mechanisms. In this perspective, drawing activities can be used to study social interaction. Joint attention, crucial in human-robot interaction, involves a robot's gazing behavior capturing interest and facilitating engagement. This thesis investigates changes in sketching behavior when individuals draw in the presence of a social child-robot observer. Therefore, two experiments were conducted. The first, to assess complexity of visually representing semantic categories, while the second, to explore joint attention in HRI and its effect on motionese.

In the first experiment, 21 adult participants were instructed to graphically represent several object categories while we were assessing their cognitive load during the task and their subjective evaluation of the task after the drawing completion. This first study allowed us to experimentally validate a construct about "complexity" that we designed for drawing tasks and that has been subsequently used in the second experiment. The experiments revealed that drawings of objects from more complex semantic categories were associated with higher cognitive load indicators, such as longer latency times, increased total drawing times, and a greater number of strokes. Clustering analysis identified three distinct clusters of drawing behaviors: Cluster 1 with high latency and total times indicating moderate challenge, Cluster 2 with low latency and total times but many strokes indicating efficient, detailed execution, and Cluster 3 with moderate latency and lower total times indicating simpler tasks.

In the second study, 53 adult participants were instructed to draw various semantic categories (e.g., a duck, an ambulance) for a child-like robot. The within-subjects experiment first presented the robot in a picture (individual condition) and later had the robot co-present and engage in joint attention behaviors with the participants (robot condition). The data collection was carried out in two sites, in Italy (N=26) and Slovakia (N=27), with two different robots, iCub and Nico. Participants significantly changed their drawing strategy in the presence of the robots by enlarging their sketches while speeding up their drawing. According to the IOS scale, the phenomenon was more evident the more the individuals perceived the robot as closer to them. The results were highly consistent between the two sites and showed

that participants put more effort into drawing understandably when a robot actively attends to their behavior. Higher clarity is obtained with increased figure size rather than simplifying the drawing. Participants did not slow down their tracing to be more comprehensible but became faster in front of the robot, possibly induced by the pressure of being observed by it.

Keywords: joint attention, triadic eye gaze, human-robot interaction, mental representation, graphic illustration

Abstrakt

V práci sa navrhuje skúmanie kresliarskych aktivít v interaktívnom kontexte s cieľom objasniť súvislosti medzi sociálno-kognitívnymi a vizuálnymi reprezentačnými mechanizmami. Vytváranie mentálnych reprezentácií, ktoré je kľúčovou oblasťou kognitívnych vied, možno skúmať prostredníctvom grafických ilustrácií, ako sú náčrty, ktoré sú spojené s ľudským poznávaním, zahŕňajúcim vnímanie, pamäť, motorické funkcie a sociálne mechanizmy. Z tohto hľadiska možno kresliarske činnosti využiť na štúdium sociálnej interakcie. Spoločná pozornosť, ktorá má kľúčový význam v interakcii človek-robot, zahŕňa správanie robota, ktorý sa pozerá, zachytáva záujem a uľahčuje zapojenie. V tejto práci sa skúmajú zmeny v správaní pri kreslení, keď jednotlivci kreslia v prítomnosti sociálneho pozorovateľa dieťaťa-robota. Preto sa uskutočnili dva experimenty. Prvý na posúdenie zložitosti vizuálnej reprezentácie sémantických kategórií, zatiaľ čo druhý na preskúmanie spoločnej pozornosti v HRI a jej vplyvu na motionese.

V prvom experimente dostalo 21 dospelých účastníkov pokyn graficky znázorniť niekoľko kategórií predmetov, pričom sme hodnotili ich kognitívnu záťaž počas úlohy a ich subjektívne hodnotenie úlohy po dokončení kresby. Táto prvá štúdia nám umožnila experimentálne overiť konštrukt o "zložitosti", ktorý sme navrhli pre úlohy kreslenia a ktorý bol následne použitý v druhom experimente. Experimenty odhalili, že kreslenie objektov zo zložitejších sémantických kategórií bolo spojené s vyššími ukazovateľmi kognitívnej záťaže, ako sú dlhšie časy latencie, dlhší celkový čas kreslenia a väčší počet ťahov. Analýza zhlukovania identifikovala tri odlišné zhluky správania pri kreslení: Zhluk 1 s vysokou latenciou a celkovými časmi, ktoré naznačujú stredne náročné úlohy, zhluk 2 s nízkou latenciou a celkovými časmi, ale s veľkým počtom ťahov, ktoré naznačujú efektívne a podrobné vykonávanie, a zhluk 3 s miernou latenciou a nižšími celkovými časmi, ktoré naznačujú jednoduchšie úlohy.

V druhej štúdii malo 53 dospelých účastníkov nakresliť rôzne sémantické kategórie (napr. kačica, sanitka) pre robota podobného dieťaťu. V rámci experimentu v rámci subjektov sa robot najprv prezentoval na obrázku (individuálna podmienka) a neskôr sa robot spolu s účastníkmi prezentoval a zapájal do spoločného správania pozornosti (podmienka robota). Zber údajov sa uskutočnil na dvoch miestach, v Taliansku (N=26) a na Slovensku (N=27), s dvoma rôznymi robotmi, iCub a Nico. Účastníci v prítomnosti robotov výrazne zmenili svoju stratégiu kreslenia tým, že zväčšili svoje náčrty a zároveň zrýchlili kresle-

nie. Podľa škály IOS bol tento jav tým zreteľnejší, čím viac jednotlivci vnímali robota ako im bližšieho. Výsledky boli veľmi konzistentné medzi oboma pracoviskami a ukázali, že účastníci vynakladajú viac úsilia na zrozumiteľné kreslenie, keď sa robot aktívne venuje ich správaniu. Vyššia zrozumiteľnosť sa dosiahla skôr pri zväčšení veľkosti postavy ako pri zjednodušení kresby. Účastníci nespomalili svoje kreslenie, aby bolo zrozumiteľnejšie, ale pred robotom sa zrýchlili, čo bolo pravdepodobne vyvolané tlakom, že ich robot pozoruje.

Kľúčové slová: spoločná pozornosť, trojitý pohľad očí, interakcia človek-robot, mentálna reprezentácia, grafická ilustrácia

Translated with DeepL.com (free version)

Contents

1 Introduction

The creation of mental representations is a deeply explored phenomenon within cognitive science. Graphic illustration through sketches can serve as a window into the internal, private world of these mental representations. The inherent complexity of drawing activities is intricately connected to human cognition, encompassing perception, memory, motor functions, and social mechanisms [1, 2]. Since drawing is a knowledge-generating activity that integrates perception, action, and cognition, previous studies implied it to explore social interaction. Joint attention, the co-orientation established in triadic interactions among the self, another agent, and a third element, plays a crucial role in human-robot interaction. Remarkably, the robot's gazing behavior effectively captures the partners' interest, facilitating the establishment of joint attention (for a review on robots and joint attention, see [3]). However, it remains unclear whether and how an individual's sketching behavior might change when drawing for an observer. Thus, we explored for the first time if and how these changes occur when individuals draw in the presence of a social child-robot observer. Humanoid robots provide an embodied context to investigate human representational mechanisms and strategies within social interaction in a controllable and repeatable manner [4].

To address these questions, two experiments were conducted. The first experiment examined the relationship between mental representations and graphic visualization, aiming to establish the construct of complexity in drawing and to test stimuli based on this construct. The second experiment focused on mental representation and social phenomena, investigating the effect of joint attention during drawing activity in a human-robot interaction framework. Specifically, we studied mechanisms related to the phenomenon of motionese in the context of drawing, evaluating whether and how human drawing style is modified when demonstrating object categories to a child-like robot observer. For this experiment, we conducted a two-site data collection using two child-like robots (iCub and Nico) to examine the influence of the robot's presence and joint attention behavior on participants tasked with drawing for them.

The thesis begins with a general introduction to semantic representations, explaining their main characteristics and hypothesized theories. We then describe the use of graphic illustration to communicate semantic representations in social contexts, discussing cognitive load theories and identifying cognitive load indicators suitable for addressing the complexity of representing semantic categories. Further, we explore the joint attention phenomenon and

its prevalent use in human-robot interaction to understand social behavior. We also examine the motionese phenomenon in social communication and its relationship with drawing activities.

The theoretical background is followed by an outline of the current research. The core of the thesis details the specific aspects of the methods used and the subsequent analysis. The results of this study will contribute to the main project, providing control group data to be compared with the experimental group for further insights.

2 Theoretical Background

2.1 Mental representations of semantic knowledge

From the perspective of cognitive psychology, the concept of mental representations involves the encoding of information in memory[5, 6], which manifests in various forms depending on the type of information being stored[7]. The main distinction lies between declarative knowledge and procedural knowledge [8]. Declarative knowledge concerns everyday life experiences' referents, such as persons, objects, events, social issues, and relations among them[8, 9]. Procedural knowledge refers to the sequence of actions performed to achieve a specific objective, such as driving a car or making an inference based on relevant facts[8, 9]. This multifaceted nature of mental representations is highlighted by recent sensory/functional theories, which assert that an object concept comprises diverse types of information derived from both acquired knowledge and direct sensory/motor experiences [10]. These experiences are processed in distinct brain regions, suggesting a highly specialized and distributed neural basis for mental representations.

One of the critical aspects of these theories is the differentiation in how sensory and abstract information is encoded. Sensory experiences tend to be represented in formats specific to the sensory modality involved, such as visual or auditory formats, whereas abstract concepts are often encoded more symbolically or semantically [7]. For instance, mnemonic representations of visual objects are situated in the ventral processing stream of visual perception, reflecting the close relationship between sensory experience and memory encoding [11]. This modality-specific encoding underscores the brain's ability to store information about an object's salient properties, such as appearance, movement, and use, within the sensory and motor systems active during the initial acquisition of that information [12].

The process of encoding also extends to how people comprehend events they observe or read about [8]. According to the literature, individuals often construct mental simulations of these events, which can include nonverbal components [8]. Such simulations do not necessarily require linguistic coding, especially when the events are directly experienced or observed [8]. In these cases, the mental model resembles the experiential form of the event. However, when dealing with verbal descriptions, the encoding process involves both linguistic and nonverbal representations[8]. This dual coding is partly because the referents of verbal descriptions must be identified to construct an accurate mental simulation, necessi-

tating a linguistic encoding of the information.

Semantic memory is a specific kind of declarative knowledge, whose impairments manifest, for example, in deficits in recognizing or using common objects and in impairments of language comprehension and production [13]. The multiple semantic systems hypothesis further elucidates how object concepts are represented within the brain [10]. According to this framework, different semantic categories are supported by various types of knowledge, such as color for fruits and both color and motion for animals[10]. These categories are represented in modality-specific semantic subsystems, each specialized for processing critical attributes of the object. This hypothesis supports the idea that our understanding and association of objects are flexible, facilitated by two types of semantic processes: taxonomic and thematic associations[10]. Taxonomic associations are based on shared properties or categories, such as grouping all fruits, while thematic associations involve context or function-based relationships, such as linking a knife with bread due to their use in a common activity

Overall, the intricate interplay between sensory experiences, abstract knowledge, and modality-specific processing underscores the complexity of mental representations. The brain's ability to encode and retrieve information in a modality-specific manner highlights the adaptability and specificity of our cognitive processes. This understanding has profound implications for social behavior. Memory encompasses a network of functions, including encoding, storing, and retrieving information (Squire, 2009). Communicating the content of mental representations primarily involves recollection, which is the process of retrieving stored information and bringing it into conscious awareness. During recollection, knowledge or experiences are voluntarily recalled by reactivating their neural representations in the association cortex.[11].

As we saw mental representations of semantic knowledge encompass a diverse array of information encoded in memory, varying based on the nature of the information—whether sensory or abstract, declarative or procedural. These representations are processed in distinct and specialized brain regions, highlighting the modality-specific nature of encoding. Sensory experiences are often encoded in formats related to the sensory modality involved, while abstract concepts are encoded more symbolically or semantically. This intricate encoding process is also evident in how individuals comprehend and simulate events, whether observed or described verbally, involving both linguistic and nonverbal components. The multiple semantic systems hypothesis further illustrates how different semantic categories

are supported by various types of knowledge, reflecting the brain's flexibility and specificity in processing object concepts.

Concluding, the complexity of mental representations underscores the sophisticated interplay between sensory experiences, abstract knowledge, and modality-specific processing. This adaptability and specificity in encoding and retrieving information reveal the profound intricacies of our cognitive processes, significantly influencing how we understand, remember, and interact with the world around us.

2.2 Graphic illustration

One potential approach to explore mental representation is through graphic illustration. Previous studies emphasized how our cultural familiarity with images, along with technological advancements, has broadened the use of visual information in both personal and professional settings [14]. Although digital tools have made it easier to create and share images, these practices are rooted in pre-digital communication methods [14]. Therefore, research has consistently advocated for the use of graphic illustration as a valuable tool in the study of cognition. Taylor et al. [15] proposed that drawing not only leaves a physical trace but also illuminates the thinking process, with images or marks serving as visible manifestations of decisions, indecisions, and verbalized thoughts. Sober suggested that pictorial representational systems can be seen as a simplified form of linguistic representation [16]. Furthermore, the 'touch' or imprint of a mark provides clues about its speed of creation [15]. Sketching is commonly described in design literature as the process of creating quick, rough visual representations [14]. Sketches represent a fundamental and intuitive means of visual expression and communication [1]. With just a few strokes, sketches encapsulate the essence of visual entities, facilitating the communication of abstract visual concepts[17].

This theory was further reinforced by recent studies which underscored the potential of graphic illustration to render the otherwise invisible contents of mental processes visible [1]. They highlighted the intricate interplay of perception, memory, and social inference during drawing production, asserting that this balance of factors enables drawings to serve as a valuable research tool for probing various cognitive phenomena [1, 2]. For example, previous studies emphasized image-making as a collaborative and communicative process. These studies are motivated by the widespread use of images, which reflects a belief in the human ability to bridge communication gaps and reach diverse audiences [14]. Previous research

suggested a qualitative and quantitative fine-grained analysis of strokes is a valid method to explore 1) how designers tend to deal with representations that are not theirs, 2) what main graphical key features constitute the inner nature of the shared information and, 3) how and when can this graphic essence be shared with collaborators [18]. Remarkably, studies on the creation and use of drawings in collaborative processes often come from the fields of Computer-Supported Cooperative Work (CSCW), Human-Computer Interaction (HCI), and Information Visualization (InfoVis)[14].

Furthermore, the significance of task design and context in shaping how drawings are utilized to convey visual ideas has been a subject of debate in prior studies [19]. In their research, Snodgrass et al. highlighted the importance of agreement between an object to be drawn and its concept [20]. They noted that the relationship between objects and their concepts is not always straightforward and can be ambiguous and subjective [20]. To address this, the researchers identified cases where such ambiguities occur and established criteria to avoid them [20]. First, subjects name the pictures to determine the picture's most common name and the degree to which subjects agreed on the name [20]. Second, subjects rated the degree of agreement between their mental image of a concept and its name by presenting subjects with the most common name of each picture [20]. They found sources of naming variance, and mean familiarity, but also complexity of exemplars differed significantly across the set of categories investigated [20].

Moreover, an empirical investigation into the recognizability of drawings revealed that label-cued category drawings were the most identifiable at the basic level, contrasting with photo-cued exemplar drawings, which exhibited the lowest recognizability [21]. Rosch et al. defined basic categories as those which carry the most information, possess the highest category cue validity, and are, thus, the most differentiated from one another [22]. They reported basic objects are shown to be the most inclusive categories for which a concrete image of the category as a whole can be formed, to be the first categorizations made during perception of the environment, to be the earliest categories sorted and earliest named by children, and to be the categories most codable, most coded, and most necessary in language [22]. Elements from different basic categories have less attribute in common, different shapes [22]. On the contrary, the members of the basic objects categories (a) possess significant numbers of attributes in common, (b) have motor programs that are similar to one another, (c) have similar shapes, and (d) can be identified from averaged shapes of members of the class [22]. This

because the categorizations that humans make of the concrete world are not arbitrary but highly structured [22]. Within the taxonomies of concrete objects, there is a specific level of abstraction where the most fundamental category distinctions are made [22]. Further, it has been shown that different representation styles follow if humans are asked to draw a specific exemplar or the general category of a given object [21].

Eventually, when we convey mental representation via graphic illustration, we should consider expectation and prediction in visual configurations [23]. Notably, graphic representations also evolve and are gradually modified if the same concept is represented more times within the same social relationship [24]. Previous research findings suggested the information value of an event represents the degree to which one's ignorance or uncertainty is reduced by the occurrence of that event [23]. In order to estimate it one must know, or assume, what the individual considers to be the probability of the occurrence of the event before it occurs [23]. Therefore, when we convey a mental representation via drawing, a way to minimize prediction error related to the delivery of the semantic content could be to choose only a member per different basic categories [22], control on the familiarity and semantic ambiguity of the category [20] or drawing within interactions with the same individuals [24]. In fact, viewers became better at recognizing objects that were drawn more frequently, while they struggled with sequences of drawings by different individuals [24]. This because sketchers rely more on shared experiences with specific viewers, leading to progressively simpler drawings that emphasized the most diagnostic features of the target object [24]. To sum up while visual resemblance is fundamental to pictorial meaning, shared experiences lead to depictions whose meanings are increasingly influenced by interaction history rather than just visual properties [24].

In conclusion, graphic illustration offers a powerful and insightful approach to exploring mental representation. The convergence of cultural familiarity with images and advancements in digital technology has significantly enhanced the creation and dissemination of visual information. As a practice deeply rooted in historical communication methods, graphic illustration is a critical tool in cognitive research, revealing the intricate thought processes through visible marks and drawings. The ability of sketches to encapsulate and communicate abstract visual concepts quickly and intuitively underscores their fundamental role in visual expression and collaborative endeavors. Empirical studies further highlight the importance of task context and the dynamic nature of the graphic representation, which evolves with

repeated use and social interaction. Thus, the utility of graphic illustration extends beyond mere depiction, providing a rich, multifaceted lens through which to examine and understand cognitive phenomena

2.3 Complexity of representing semantic categories

Notably, graphic illustration represents a potential tool to study cognitive performance. However, literature to assess the cognitive load implied in representing specific drawing objects is scarce. The stimuli validation experiment of this study aims to evaluate how the cognitive load variance across representing object categories to assess the task complexity and control over the confounding variable.

2.3.1 Cognitive Load

Cognitive Load Theory is based on a cognitive architecture that consists of a working memory limited in capacity and time when it comes to holding or processing novel information [25]. According to the triarchic theory of cognitive load based on CLT, there are three kinds of cognitive processing during learning that can contribute to cognitive load [26]. Intrinsic load reflects the nature of learning materials and is positively related to the number of interactive elements of learning materials [27]. Additionally, it cannot be manipulated by the design of the task (e.g., the calculation of 2 + 2, versus solving a differential equation)[28]. Extraneous load, imposed by suboptimal instructional design, depends on the number of interactive elements that are present not because of the nature of the information but because of the way the information is presented [29]. Germane load refers to the actual working memory resources allocated to deal with intrinsic cognitive load [30, 31, 29].

In the traditional perspective, cognitive load was defined as the burden a particular task imposes on the cognitive system. It was conceptualized in two dimensions: a task-based dimension, where mental load reflects task demands, and a learner-based dimension, where mental effort represents the cognitive resources allocated to meet these demands [32]. However, this theory has evolved over the years, sparking debate about the different interpretations of the components underlying cognitive load [25].

Consequently, research has often used effort investment and task difficulty interchangeably to assess cognitive load [25]. Although the traditional Cognitive Load Theory (CLT) perspective distinguishes between intrinsic and germane load, there is no consensus on a

method to separately assess these constructs while preserving the predictive value of cognitive load indicators [30]. The challenge in measuring cognitive load lies in identifying reliable cognitive load indicators[33]. Therefore, this study aimed to identify specific cognitive load indicators. To do this we identify some popular cognitive load indicators to test the average sensitivity to task demand. The following sections will illustrate the cognitive load indicators selected and adapted for the current study.

2.3.2 Task Complexity

Task complexity is defined by the number of interacting elements required to solve the task [34]. This differs from task difficulty, which is typically defined by the mean probability of solving the task [34]. Task complexity influences cognitive processing; as task complexity increases, more cognitive resources are required [34]. Previous studies have examined the impact of item-level complexity and found that item complexity serves as a cue for perceived difficulty [33]. Thus, the number of interacting elements required to solve a task reveals perceived difficulty[33], reflecting intrinsic task complexity[34].

In most research on cognitive load, learning outcomes are measured by different tasks, varying in complexity[34]. The internal complexity of the learning materials is measured by the degree of interconnectedness between essential elements of information that should be considered in working memory at the same time (element interactivity) [1] [27]. During intrinsic processing, the learner engages in cognitive processing that is essential for comprehending the material and that depends on the complexity of the material, namely the number of interacting elements. However, even if the total number of elements may be relatively small, the interactivity of the elements renders the tasks difficult [35]. Accordingly, even if the number of elements in a task stays the same, changes in how these elements interact can affect the task's complexity[29]. Thus, the higher the number of interacting information elements a task contains, the more difficult it is and the higher the intrinsic load it imposes on working memory [36].

Importantly, the interactivity among elements is directly linked to a learner's level of expertise [35, 29]. As such, by varying the learners' expertise, one can indirectly modify the levels of cognitive load, through the alteration of element interactivity levels [35]. Previous studies reported that experts and novices perceive the complexity of a task differently due to

[1] The term "interactivity," as defined by Sweller (2010), refers to how elements interact with one another.

their differing levels of familiarity and understanding. An expert can recall and integrate the components of a problem, such as an equation and its solution, from long-term memory as a single unit, making the task seem simpler (low element interactivity)[35, 29]. In contrast, a novice sees these components as separate and complex (high element interactivity)[35, 29]. Therefore, the intrinsic load imposed by a task consists of the inherent complexity of the content (i.e., interacting information elements) concerning the learner's level of expertise [36].

Moreover, the complexity posed by a task can be mitigated by "component redundancy," which refers to the overlap in demands from different parts of the task[37]. If knowledge or skills needed for one part of the task apply to other parts, the overall knowledge or skill requirement decreases[37]. An extreme example of total redundancy is a task where the same action must be repeatedly performed, thus simplifying the task[37]. The concept of redundancy also applies equally well to tasks that are either predominantly cognitive or predominantly motor-physical type tasks[37]. These findings demonstrated that both individuals' levels of expertise and task redundancy are confounding variables to account for when measuring task complexity. Thus, a simple and general index for task complexity is the unit-weighted summation of distinct task interacting elements needed to solve the task, where an interactive element is considered distinct if it is non-redundant with other elements[37, 34].

2.3.3 Subjective rating scales

Another widely used method in CLT research for assessing cognitive load is subjective rating scales [34, 33]. These ratings take the form of metacognitive judgments, such as ease of learning, judgments of learning, feelings of rightness, or confidence [33]. For consistency with CLT terminology, it is important to note that these judgments are meant as appraisals[38, 33]. However, Scheiter et al. (2020) further conceptualized various Cognitive Load Theory (CLT) appraisal items from a metacognitive perspective, arguing that the phrasing of effort investment items can reflect both motivational and cognitive aspects related to processing and task performance [33]. A recent study by Hoch et al. [33] reviewed appraisal methods, finding that load-related appraisals may rely on external cues. They reported these appraisals express sensitivity to task demands but are often dissociated from objective measures of success [33]. Specifically, the effort the task demands can be referred to as a data-driven effort [39]. It is based on bottom-up processing, focusing on task characteristics that learners

cannot control, similar to asking about the difficulty of the task. The scales typically include a single question, such as "Please rate the amount of mental effort invested in the task," with responses ranging from "very very low mental effort" to "very very high mental effort," as seen in the widely used mental effort scale by Paas (1992)[33]. Appraisals are often collected using 5-, 7-, or 9-point Likert scales[33].

2.3.4 Behavioral data

According to the time-based resource-sharing model, the cognitive load of a task depends on the proportion of time it captures attention, thereby hindering other attention-demanding processes [40]. Accordingly, previous studies hypothesized that the disruptive effect on the concurrent maintenance of memory retrievals and response selections increases with the duration of these processes [40]. These findings indicate that working memory operates sequentially and in a time-based manner, where both processing and storage depend on a single, general-purpose attentional resource [40]. This resource is essential for executing processes that construct, maintain, and modify temporary representations [40]. Hoch et al. examined the underlying bases and the predictive value of mental effort, task difficulty, and metacognitive confidence appraisals in three cognitively demanding problem-solving tasks by using metacognitive concepts, paradigms, and measures. The results suggested that using response time as a cue can be informative for mental effort [33]. Baars et al. have shown that invested effort is related to time investment [33]. Specifically, the effort an individual chooses to invest is termed goal-driven effort, which relies on top-down processing and reflects voluntary decisions made by learners [33]. Several behavioral parameters can serve as an indicator of cognitive load such as invested time-on-task and response time. [34].

Time-on-task. Results of a series of correlational analyses and repeated-measures ANOVA showed that time-on-task, which can further be divided into time-on-planning and time-on-speech, proved to be a valid measure of cognitive load [41]. All cognitive processes take time. The amount of time needed to reach a solution is affected by several factors, including the complexity of the task, the learners' prior knowledge, the time needed to search for information, and so forth. Nevertheless, time-on-task is directly related to cognitive processing and in a good experimental setting, it is possible to control for most of these factors by measuring them or by randomly assigning participants to different conditions. In this way, it is possible to come closer to the basic processing variables causing differences in time-on-task.

The response time. The abilities to select appropriate responses and suppress unwanted actions are key executive functions that enable flexible and goal-directed behavior [42]. Response time refers to the time an agent needs to make a decision [43]. Therefore, the efficiency of processing can be measured with variations in the effort required by the participants [44]. Tests performed on response time distributions are proving to be useful tools in determining the workload capacity (as well as other properties) of cognitive systems [44]. Accordingly, the detection response task (DRT) is a well-validated method for measuring cognitive workload that has been used extensively in applied tasks, for example, to investigate the effects of phone usage or passenger conversation on driving, but has been used sparingly outside of this field [45].

However, the literature suggested that none of the cognitive load indicators could be considered a reliable method if taken separately from the others. In these cases, the risk is the collected data may not have predictive value over the research question. First, metacognitive research has systematically shown that appraisals are prone to biases, as they are based on heuristic cues and lay theories [38, 33]. Similarly, recent studies encourage considering multiple sources for actual and perceived difficulty beyond response time, even when response time as a cue has been shown to result in monitoring biases [33]. Eventually, research has demonstrated that perceived mental effort and task difficulty are significantly higher when measured at the end of the learning phase (delayed) compared to the average ratings collected immediately after each learning task [46]. Therefore, subjective ratings as the time data should be recorded immediately after each task, in which case they presumably reflect the accumulated cognitive load [46].

2.4 The Significance of Joint Attention in Social Cognition and Human-Robot Interaction

Previous studies have debated the importance of studying mental representations and social factors [8]. Joint attention refers to moments when a child and adult are focused on the same thing, but for most researchers, it also includes the notion that the participants are both aware that the focus of attention is shared [47]; that is, joint attention involves the child and adult coordinating mutual engagement with their mutual focus on a third entity [48].

The term joint attention has also been used to refer to a whole complex of supposedly 'social cognitive' behaviors that emerge toward the end of the first year of life (e.g., social

referencing, pointing, and gaze following). The social-cognitive model of joint attention proposes that social cognition is necessary for the development of functional joint attention in infancy [48, 49]. As the integrative anterior-posterior capacity to attend to overt or covert information about self and others matures with advances in representational abilities, a fully functional, adaptive, human social-cognitive system emerges [49]. Around 9 months, children begin to recognize when their attention is aligned with their caregiver on an object and begin pointing things out to align their attention [48, 50]. Joint attention transforms children's interactions with others [26]. It has been shown to support word learning [50], to support our ability to cooperate with others [50], and to play so prominent a role in social interaction that a poor ability to engage in joint attention predicts pervasive social difficulties [49].

Joint attention is strictly connected to gaze movement. Functionally, starting between 4 and 6 months of age, the anterior attention system integrates the internal monitoring of one's own control of gaze direction, and its relations to goal-directed behavior, with external monitoring of the relations between others' gaze direction and their behavior [49]. Here's a breakdown of some of the key skills related to eye gaze.

- Eye contact: orienting visual attention to another person.
- Gaze shifting: disengaging from focusing on one person, object, or event to attend to something new.
- Triadic eye gaze: three-way gaze shifting from a person to an object, event, or other person and back.

Previous research explained joint action as some coordination mechanisms depending on sensorimotor information shared between co-actors, thereby making joint attention, prediction, non-verbal communication, or emotional states possible [51]. Others' eye movements are an important source of information about what others see and about their internal states [48]. Joint attention relies on co-actors' ability to monitor each other's gaze and attentional states [51]. Sharing gaze affects object processing by making attended objects motorically and emotionally more relevant [51]. Together, these findings demonstrate the important role of gaze information in joint attention.

Notably, joint attention played a crucial role in human-robot interaction. Previous studies reported the gazing behavior of the robot effectively gained the partners' interest and drew

them to establish joint attention (for a review on robots and joint attention, see [3]). Previous studies demonstrated fundamental competencies for social robots, the ability to contextualize and portray joint attention (joint attention) and knowledge in the context of human-robot interaction (HRI)[52]. This proficiency is essential for achieving shared objectives through joint action[53, 54] and may constitute the basis for establishing a relationship between humans and robots.

2.4.1 Motionese

When we communicate to others we share our representation of the world. However, the receiver can influence the way we express our thoughts. Motionese phenomena are action style modifications, such as slowing down one's movements, introducing more segmentation, and standing closer when demonstrating a desired behavior, that occurs naturally when human caregivers interact with infants [55]. Several studies support the idea that humans use a similar movement style when interacting with robots [56, 57, 58, 59]. In this context, the learner's behavior can also influence how a tutor demonstrates an action. Nagai et al. [57] showed that humans not only modify their behavior when demonstrating action to a robot with exaggeration in space and synchronization in time, but the robot's bottom-up attention also influenced motions used by human teachers.

2.5 The current study

2.5.1 General objective

Graphic illustration via sketches can be considered a gateway to the internal, private world of mental representations. The inherent complexity of drawing activity is connected to human cognition through perception, memory, motor, and social mechanisms [1, 2].

Humanoid robots can provide the embodied context to investigate human representational mechanisms and strategies within social interaction in a controllable and repeatable way [4]. Yet, whether and how one would modify its sketching behavior when asked to draw for an observer is still unknown. For this reason, we investigated for the first time, if and how these changes occurred in front of and for a social (child) robot observing drawers in their task. This study aims to explore the sensitivity to task difficulty related to the stimuli during the drawing semantic categories task. Therefore, since the literature does not

provide a specific procedure to assess cognitive load on drawing tasks, the motivation is to identify and validate a method to test stimuli in such an experimental context. Furthermore, this work investigates whether similar joint attention phenomena occur during the activity of drawing for a child robot observer and whether the robot's physical presence influences this interaction. Moreover, we aimed to study mechanisms related to motionese in the context of drawing to evaluate whether and how the human drawing style is modified while graphically demonstrating object categories to a robot observer. By examining these dynamics, we hope to understand better how drawing activities can facilitate joint attention and enhance collaborative processes involving robots.

2.5.2 Methodology

Therefore two experiments have been designed. The first one was motivated by testing the complexity of representing semantic categories. To do this we tested the complexity of drawing categories construct to assess the task difficulty.

Further, we carried out a two-site experiment by using two child-like robots (iCub and Nico) to investigate the influence of the robot's presence and joint attention behavior on participants who were asked to draw for them. In particular,

2.5.3 Hypotheses

For the first experiment we expect that the drawings of objects from more complex semantic categories (e.g., abstract concepts, intricate shapes) will result in higher cognitive load indicators, such as longer latency times, total drawing times, and a greater number of strokes, compared to drawings of objects from simpler semantic categories (e.g., basic shapes, concrete objects)

For the second experimend: 1) we expect the eye gazing of the child-robot effectively gain joint attention and affects participants' perception of closeness to the robot. 2) we expect joint attention to influence communication strategies during drawing categories for a child-robot observer.

3 Stimuli validation experiment

In the previous section, various techniques for assessing cognitive load during task performance were discussed. While these techniques have shown effectiveness across different learning tasks, they may not be perfectly suited to the demands of specific complex environments, such as representing object categories via graphic illustration. Hence, we have identified options for adapting these techniques to the current experiment. This adaptation involves combining existing methods, such as rating scales and behavioral data. Inspired by previous methods [33], we collected the results bottom-up process via self-appraisal and top-down with a time stamp to avoid bias. Additionally, we considered the level of element interactivity. Therefore, we wanted to test whether the methods suggested in literature can be adapted to assess cognitive load on representing diverse semantic categories within a drawing task. We expected the results to reveal significant correlation across subjective ratings, behavioral data and element interactivity, supporting the complexity of representing semantic categories via drawing construct. Particularly, we expected participants to enjoy less likely the activity and like the outcome of drawings to a lesser extent when they were perceived as difficult.

3.1 Methods

3.1.1 Participants

A sample of N = 21 participants (M 10 W 11 NB 0) underwent the experimental session at the Italian Institute of Technology (IIT, IT). The sample consisted of 21 Italians. All participants provided written informed consent before their participation. The study received approval from the regional ethical committee, Comitato Etico Regione Liguria.

3.1.2 Design

In this within-subject experiment, participants took part in a drawing task to rank the complexity of drawing representations. The drawing task consisted of 17 trials, one per drawing category. Therefore, participants draw a different object per trial. As suggested in the theory (see, Section **??** we asked for ratings and collected time data after each of the 17 drawing sessions.

3.1.3 Setup

The experimental room had a laptop and a touchscreen for the drawing activity. The model used for the touchscreen was the ELO 2002L (436.9x240.7 mm, 1920x1080 px, 60 Hz). We used Python 3.8 and Psycopy (version 2023.1.2) to create the virtual environment and a suitable graphic interface to make participants draw and answer questions while collecting all the data.

3.1.4 Behavioral measurements and Variables

For each drawing, we determined the variables to be assessed. Participants' cognitive load was measured via self-report scales, time data (response time and task completion time), and interactivity to a secondary monitoring task, see section 2.3.1.

Latency time was defined as the duration between the stimulus presentation (the time the category name was displayed on the screen) and the response time (the time participants touched the screen to draw the first stroke). This metric served as an indicator of workload. A longer latency time suggests a more challenging effort in devising a strategy (see Section 2.3.4).

Drawing time . It is the total time of actual drawing activity, computed as the total time drawers use to sketch the strokes. It is the difference between the final timestamp of a stroke and its initial one. It captures the proportion of time the task captures the participant's attention. Therefore, it is informative about the time and the attentive resources voluntarily invested. A longer drawing time suggests a higher investment of *mental effort* (see Section 2.3.4).

The total number of strokes . It is the total number of strokes used to complete the drawing. Every time participants lifted their fingers and touched the screen again, a new stroke was created. It can indicate the *amount of interactive details* inserted in the drawing (see Section 2.3.2).

3.1.5 Questionnaires, scales and subjective ratings

Subjective rating scales. Different techniques are available for measuring cognitive load. As mentioned in Section 2.3.3, subjective rating is commonly used for the assessment of cogni-

tive load. Therefore, we adapted the subjective rating question to the graphic representation task. Participants rated the perceived difficulty. Besides, they rated their experience in terms of perceived enjoyment and the aesthetic value of their production in terms of perceived likability.

Perceived difficulty . It is the subjective difficulty rating assigned to each drawing. It captures the sensitivity experienced by every participant in its realization, because it relies on external cues that subjects cannot control. It is informative about the task demand (see Section 2.3.3).

Perceived enjoyability Specifically, it is the subjective value assigned to each drawing session, addressing the participant's activity appreciation level.

Perceived likability Eventually, likability represents the subjective value assigned to each drawing to depict the participant's appreciation level of the outcome of their activity.

This study assigned ratings on a scale from 1 to 7, utilizing Likert-type scale response anchors/numbers. We opted for a 7-point Likert scale anchor/number for each variable. Here, they follow an example for the three questions for the subjective rating and related scales: How difficult was it to draw the " BEE "? "(1 - not difficult, 7 - extremely difficult)" "How much did you enjoy drawing the " BEE "? "(1 - not enjoyed, 7 - extremely enjoyed)", "How much do you like your drawing of the " BEE " "(1 - not liked, 7 - liked a lot).

3.1.6 Experimental sessions

Once in the experimental room, informed consent and instructions were supplied to the participants. Instructions were explained by the experimenter, but participants were provided also with written instructions they could keep during the experimental session. At this point, the participants were allowed to start. At first, participants performed an exemplifying trial and then 17 consecutive drawings/trials, one drawing per object category. Each trial began by displaying the name of the category participants were asked to draw on the touchscreen. The drawing time started when participants communicated they were ready to draw by pressing an appropriate button on the screen. A white canvas appeared on the screen. The drawings were performed with the index finger of the dominant hand. After each drawing was completed participants were asked to answer questions about the enjoyment of performance,

likeability, and perceived difficulty of the drawing. Sessions were designed without any time limit to let participants draw without any pressure and hence avoid influencing their performance.

3.1.7 Stimuli (object categories)

18 object categories were chosen from the Google Quick Draw Dataset [2] [3] [60]. Figure 1 provides a clear schema of the stimuli. Categories were selected based on their semantic taxonomy or shape. Therefore, some categories were quite similar because they belonged to the same semantic group (transports: bus, train, ambulance; animals: sheep, lion, owl, bee, spider), whereas others because they have similar starting shapes (round: pizza, face, owl). We established no trait limits to avoid influencing participants to worry about fitting within a specific range of traits. This approach also helped us avoid increasing extraneous load (see Section 2.3.1). Selected categories also varied in complexity of the figure, and number of details. The names of the categories were sequentially presented to participants on the touchscreen as a visual stimulus, with red writing on a black background, lasting for 4 seconds each. To mitigate bias, the order of the categories was randomized and stimuli were shown displaying the object category label rather than an exemplary image. After the category name was displayed on the touchscreen, a white canvas opened as a virtual sheet on which participants had to sketch.

3.1.8 Data analysis

To assess the complexity of representing categories through sketches, we compiled a data frame containing latency time, total drawing time, total number of strokes, and perceived difficulty rankings. Raw data were then organized into a proper dataset using the Pandas and NumPy libraries. Additionally, Tukey's method was employed to identify outliers. To understand the overall trend in perceived difficulty, we calculated means and standard deviation of difficulty ranking (see Section 3.2). Furthermore, we conducted a more in-depth analysis using the Jamovi library. To explore correlations among variables, we utilized the Spearman correlation test, whereas the snowCluster module was used to perform a K-means cluster analysis to categorize the object categories into three clusters.

[2] https://github.com/googlecreativelab/quickdraw-dataset

[3] https://console.cloud.google.com/storage/browser/quickdraw_dataset/full/raw;tab=objects

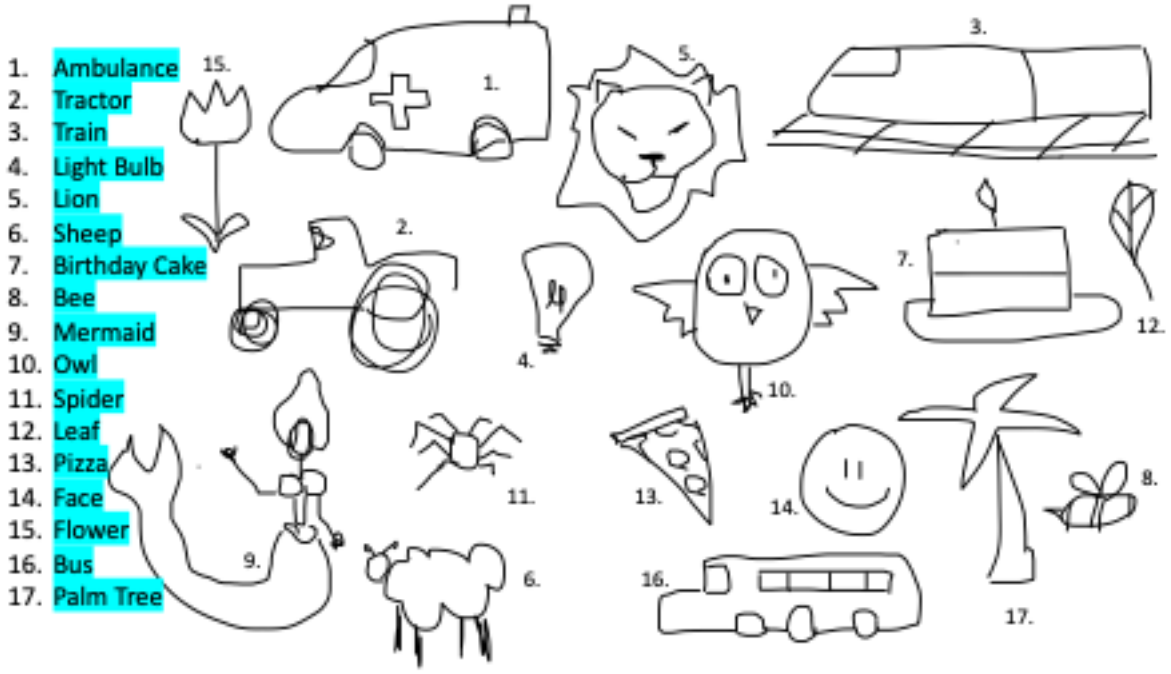

Figure 1: The 17 selected categories represent the stimuli in the stimuli validation experiment

3.2 Stimuli validation experiment results

3.2.1 Complexity ranking

The mean difficulty ratings, median values, standard deviations (SD), and Shapiro-Wilk test results for normality for each category are summarized in Table 1. For the categories "Ambulance", "Bee", "Birthday Cake", "Face", "Flower", "Leaf", "Light Bulb", "Lion", "Owl", "Palm Tree", "Pizza", "Spider", "Tractor", and "Train", Shapiro-Wilk tests revealed significant departures from normality at the $p < .05$ level. Specifically, the Shapiro-Wilk tests for these categories indicated p-values as follows: Ambulance ($p = 0.027$), Bee ($p = 0.046$), Birthday Cake ($p = 0.020$), Face ($p = 0.008$), Flower ($p < 0.001$), Leaf ($p < 0.001$), Light Bulb ($p = 0.003$), Lion ($p = 0.015$), Owl ($p = 0.035$), Palm Tree ($p = 0.047$), Pizza ($p = 0.031$), Spider ($p = 0.039$), Tractor ($p = 0.003$), and Train ($p = 0.041$).

For the "Bus" and "Mermaid" categories, the Shapiro-Wilk tests did not indicate significant departures from normality, with p-values of 0.481 and 0.268, respectively. For the "Sheep" category, while the Shapiro-Wilk test approached significance ($p = 0.079$), it did not reach conventional levels of significance. These results suggest that for most of the categories, the difficulty rankings do not follow a normal distribution. However, for "Bus" and "Mermaid," there is no evidence of departure from normality based on the Shapiro-Wilk test. Notably, no outliers emerged from the test.

Table 1: Descriptive Statistics

Class	Mean	Median	SD	Shapiro-Wilk p
Ambulance	4.95	5	1.86	0.027
Bee	3.86	4	1.53	0.046
BirthdayCake	3.33	3	1.53	0.020
Bus	4.19	4	1.66	0.481
Face	3.71	3	1.90	0.008
Flower	2.00	2	1.05	<.001
Leaf	1.86	2	1.20	<.001
LightBulb	2.52	2	1.63	0.003
Lion	5.29	5	1.23	0.015
Mermaid	4.67	5	1.20	0.268
Owl	5.14	5	1.53	0.035
PalmTree	3.38	3	1.47	0.047
Pizza	3.14	3	1.62	0.031
Sheep	4.10	4	1.73	0.079
Spider	2.95	3	1.56	0.039
Tractor	5.52	6	1.29	0.003
Train	4.33	5	1.93	0.041

3.2.2 Spearman correlation

Spearman's rho correlation coefficient was computed to estimate the monotonic relationship between the perceived difficulty and the other variables. Table 2 showed Pearson analysis results. Findings reported: a strong positive correlation between difficulty ranking and total time: Spearman's rho = 0.897, p < .001; a strong positive correlation between difficulty ranking and likability ranking: Spearman's rho = -0.940, p < .001.

A moderate positive correlation between difficulty and latency time, Spearman's rho = 0.637, p = 0.007; a moderate positive correlation between the number of strokes and the difficulty ranking, Spearman's rho = 0.667, p = 0.003.

Further, a significant positive correlation emerged from the relationship between other variables. Remarkably, results showed there was a moderate positive correlation between total time and latency time, Spearman's rho = 0.706, p = 0.002; there was a moderate positive correlation between the number of strokes and total time, Spearman's rho = 0.782, p < .001; there was a weak positive correlation between enjoyment and likability, Spearman's rho = 0.417, p = 0.096.

Moreover, findings reported there was a strong negative correlation between total time and likability, Spearman's rho = -0.829, p = < .001; there was a moderate negative correlation

Table 2: Spearman's rho correlation coefficients and p-values for the relationships between Difficulty, Latency, Total Time, Total Strokes, Enjoyment, and Likability. Significant correlations are indicated with *p < .05, **p < .01, and ***p < .001.

Spearman Correlation Matrix							
		Difficulty	Latency Time	Total Time	N Strokes	Enjoyment	Likability
Difficulty	Spearman's rho	–					
	p-value	–					
Latency Time	Spearman's rho	0.637**	–				
	p-value	0.007	–				
Total Time	Spearman's rho	0.897***	0.706**	–			
	p-value	<.001	0.002	–			
N Strokes	Spearman's rho	0.667**	0.552*	0.782***	–		
	p-value	0.003	0.022	<.001	–		
Enjoyment	Spearman's rho	-0.425	-0.496*	-0.502*	-0.389	–	
	p-value	0.089	0.043	0.040	0.123	–	
Likability	Spearman's rho	-0.940***	-0.634**	-0.829***	-0.700**	0.417	–
	p-value	<.001	0.006	<.001	0.002	0.096	–

between the number of strokes and likability, Spearman's rho = -0.700, p = 0.002; a weak negative correlation between enjoyment and total time, Spearman's rho = -0.502, p = 0.040.

3.2.3 K-mean cluster analysis

Table 3: Centroids of clusters representing three distinct trends identified in the dataset. Each column represents the average values of various metrics (Latency Time, Total Time, Number of strokes, Perceived difficulty) for the respective cluster, highlighting the central tendency of the trends within each cluster.

Centroids of clusters Table					
	Cluster No	LatencyT_avg	TotalT_avg	No Strokes_avg	Difficulty_avg
1	1.00	-0.472	-0.122	-0.129	-0.132
2	2.00	-0.522	-1.343	-1.179	-1.338
3	3.00	0.899	1.038	0.936	1.046

The Table 3 reported three different cluster representing three trends:

- Cluster 1:Latency Time (avg)= -0.472; Total Time (avg)= -0.122; Number of Strokes (avg)= -0.129; Perceived Difficulty (avg)= -0.132;
- Cluster 2: Latency Time (avg)= -0.522; Total Time (avg)= -1.343; Number of Strokes (avg)= -1.179; Perceived Difficulty (avg)= -1.338
- Cluster 3: Latency Time (avg)= 0.899; Total Time (avg)= 1.038; Number of Strokes (avg)= 0.936; Perceived Difficulty (avg)= 1.046

3.3 Complexity of representing semantic categories construct experiment discussion

The results of the Shapiro-Wilk tests revealed significant departures from normality in the difficulty ratings for the majority of the categories, indicating that the difficulty ratings for these categories do not follow a normal distribution.

The significant deviations from normality in the difficulty ratings for most categories suggest that non-parametric tests or data transformations might be more appropriate for analyzing difficulty ratings in these cases.

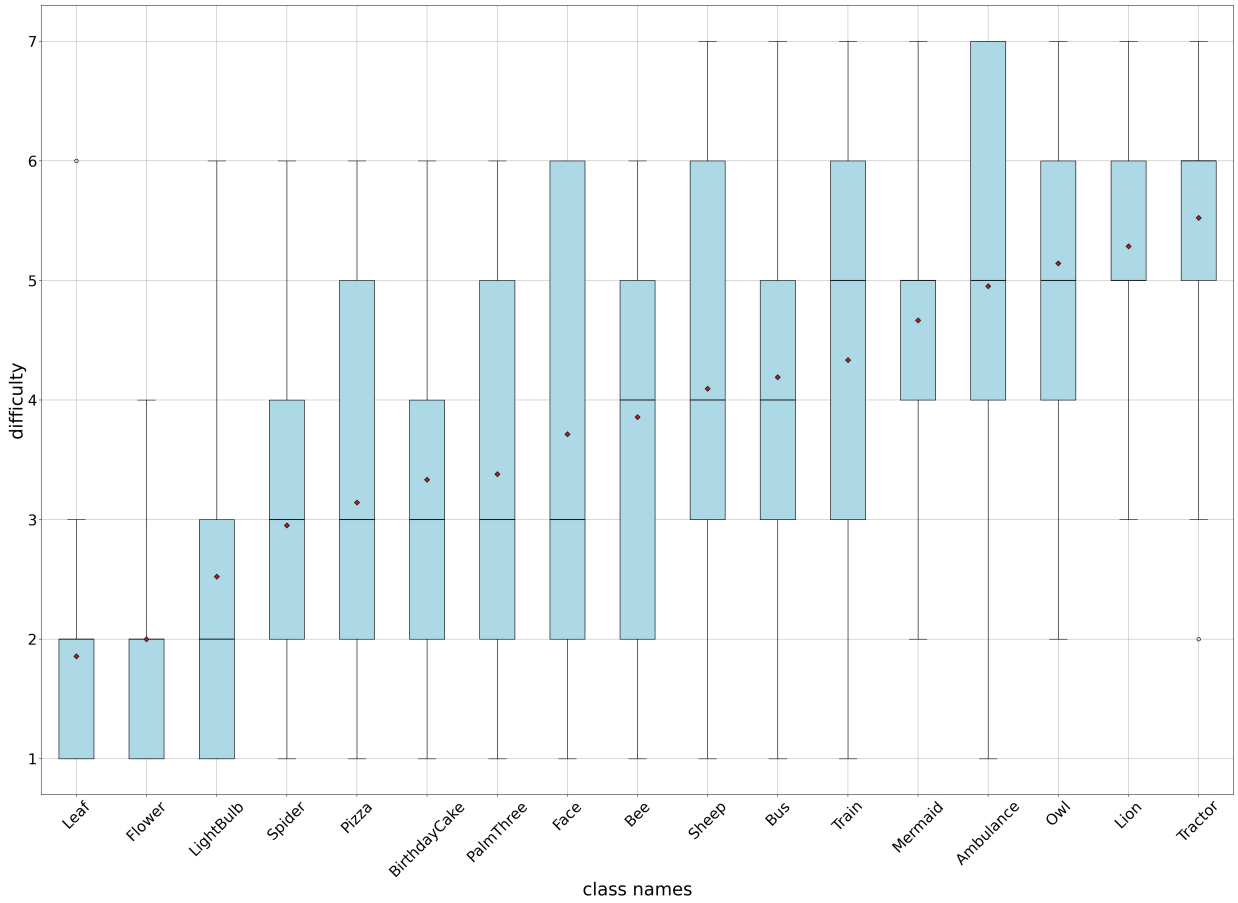

Figure 2: Comparing the reported difficulty of categories within participants (N=21). The boxes represent the interquartile range (IQR), with the horizontal line inside each box indicating the median difficulty rating. The whiskers extend to the smallest and largest values within 1.5 times the IQR from the quartiles, and the dots represent outliers.

Figure 2 shows the distribution of object categories for what concerns participants' perceived difficulty. Categories like "Leaf", "Flower", and "Light Bulb" have lower median difficulty ratings, reflected in their mean values of 1.86, 2.00, and 2.52, respectively. These categories also have relatively low standard deviations (SD), indicating less variability in

participants' ratings.

In contrast, categories such as "Lion", "Owl", and "Tractor" have higher median difficulty ratings, with means of 5.29, 5.14, and 5.52, respectively. The higher difficulty ratings are evident in the wider boxes.

The "Bus" and "Mermaid" categories have median ratings of 4 and 5, respectively, with means of 4.19 and 4.67. The Shapiro-Wilk tests for these categories were not significant ($p = 0.481$ and $p = 0.268$), suggesting that their difficulty ratings follow a more normal distribution. This is supported by the more symmetric appearance of the boxes and whiskers in the plot, indicating less skewness.

Overall, the boxplot visually confirms the statistical results from the Shapiro-Wilk tests and the descriptive statistics. Categories with significant departures from normality typically exhibit more skewed distributions and greater variability, as shown by the larger interquartile ranges. Future research could explore the underlying causes of the non-normal distributions observed in these categories. For instance, it would be useful to investigate whether certain object categories are inherently more variable in terms of perceived difficulty, and whether this variability is influenced by factors such as cultural differences, individual experiences, or the specific attributes of the objects being rated.

Furthermore, Sperman correlation analysis observed a significant correlation among the variables, with particular emphasis on the strong association between perceived difficulty and other factors (see Table 2). The findings confirmed that participants' reported difficulty (N=21) during drawing sessions correlates to latency time, indicating the sensitivity to intrinsic task difficulty is subjected to processing object identity and graphic representation strategy. The positive correlation suggested that increased reported difficulty led to longer latency times. This pattern was consistent for total time and difficulty as well. Findings suggested the more subjects were dedicating attentive resources to a drawing, the higher the sensitivity to the inherent difficulty of the category. Accordingly, the positive correlation between the number of strokes and difficulty, suggested the higher the number of elements composing the draw, the more sensitive the participants were towards the task difficulty. Conversely, likability exhibited a strong negative correlation with difficulty. This findings suggested that participants were less likely to like the outcome of the activity when they were more sensitive to intrinsic task difficulty.

Further, strong positive correlation were noted for total time, which reflects the amount

of attentive resource invested (mental effort) and the number of elements in the drawings, an index of visual complexity reflecting the intricacy of lines. Higher mental effort in representing the object categories was associated with more time spent on drawing, as evidenced by a positive correlation. Similar trends were noted for total time, which reflects the amount of attentive resource invested (mental effort) and latency time. The positive correlation suggested that the more participants allocated their attentive resource on the task the higher was the response time. Oppositely, a negative correlation characterised the relationship between likability, enjoyment and total time, suggesting the higher was the mental effort invested, the less participants enjoyed the experience and the outcome of their drawings.

The positive correlation between the latency time and the number of strokes suggested the higher was the time required to find a strategy, the higher was the visual complexity of the sketch.

Notably, enjoyment and likability exhibited a negative correlation with almost all the cognitive load indicators, suggesting that participants were less likely to enjoy the activity and like the outcome of drawings when they were more complex to represent.

Detailed correlation coefficients and significance levels are provided in the accompanying statistical analysis. For a more comprehensive understanding, have a look at Table 2.

The analysis revealed significant correlations between perceived difficulty and various tasks performance metrics, highlighting that increasing difficulty lead to longer latency times, more strokes and longer total time. These factors were associated with lower likability and enjoyment of the drawing tasks. The findings confirmed our hypotheses that higher sensitivity to the inherent difficulty of the task influence the other cognitive load indicators providing an interesting method to explore the complexity of visual representing semantic categories.

With the k-means clustering, we have identified categories of drawings with similar quality (perceived difficulty, number of strokes, drawing time, latency). See Table 3. Cluster 1 shows above-average values in all metrics, suggesting that participants in this cluster experienced higher latency, took more time to complete tasks, made more strokes, and find the tasks more difficult. Similarly, cluster 2 is characterized by below-average values across all metrics, indicating a trend where participants experienced lower latency, completed tasks in less time, with fewer strokes, and perceived the tasks as less difficult. Yet, cluster 3 has values close to the average, with slight deviations below the mean. It represents a more balanced trend where participants' experiences are near average for all metrics, with a slight tendency

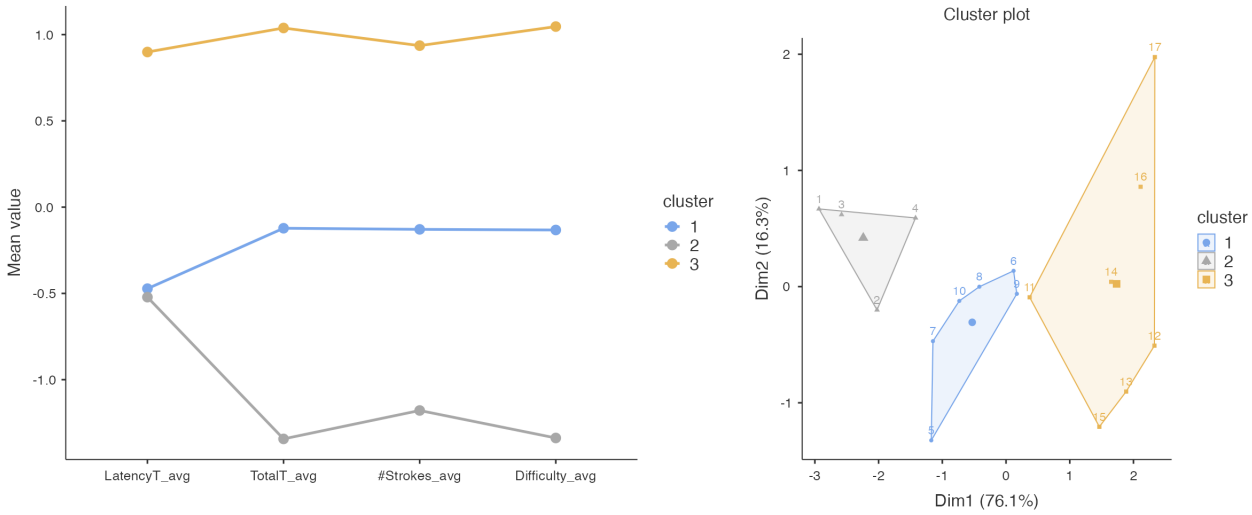

Figure 3: *Left:* graphical representation of three distinct trends based on participants' recorded latency time, total time, number of strokes, and difficulty. *Right:* plot of clustered categories with similar trends in latency time, total time, number of strokes, and difficulty.

towards lower values. The trends are displayed in the cluster plots, see Fig. 3. The plot on the left illustrates the average values of four metrics (Latency Time, Total Time, Number of Strokes, and Perceived Difficulty) for three distinct clusters. While the X-axis represents the four variables, the Y-axis shows the mean values of these variables, standardized to a range where values can be positive or negative. The cluster plot on the right (see Figure 3) shows the distribution of data points across the three clusters in a two-dimensional space defined by two principal components (Dim1 and Dim2). X-axis (Dim1) represents the first principal component, accounting for 51.2% of the total variance in the data. Y-axis (Dim2) represents the second principal component, accounting for 40.8% of the total variance in the data. Together, these two dimensions capture 92% of the total variance, providing a good representation of the data in two-dimensional space. Each cluster is represented by a different color: blue (Cluster 1), grey (Cluster 2), and orange (Cluster 3). The overall trend shows: **Latency Time** (LatencyT_avg) Notably, Cluster 1 and Cluster 2 exhibit lower-than-average latency times, indicating that participants in these groups responded more quickly. This suggests higher responsiveness and potentially a lower cognitive load demonstrating efficient task performance. This efficiency could be attributed to better strategy, higher skill levels, or lower cognitive load. Diversely, cluster 3 shows significantly higher latency times, indicating slower responses. Participants in this cluster might have faced challenges in processing a representation, reacting unpromptedly, which could be due to less efficient strategies, therefore a higher cognitive load.

Total Time (TotalT_avg) In cluster 1 the total time is close to average, indicating standard performance levels without significant deviations. Participants in this cluster likely performed tasks at a normal pace. Inversely, in cluster 2 participants took significantly less time to complete the tasks, likely due to better focus or familiarity with the tasks. Cluster 3: This cluster shows higher total time, suggesting that participants took longer to complete the tasks. This could indicate complexity perceived in the tasks, and therefore more attentive and time resource allocated to complete it.

Number of Strokes (No Strokes_avg) In cluster 1, the number of strokes for these categories is average, suggesting typical task execution without significant cognitive resources expenditure. Conversely, in cluster 2 fewer strokes were recorded for drawn categories, indicating that participants required fewer actions to complete the tasks. This efficiency can be linked to less cognitive load task-related and more effective problem-solving approaches. Eventually, cluster 3 represents categories that had more strokes, indicating that participants required more actions to complete the tasks. This could be due to less effective strategies and greater difficulty in task execution.

Perceived Difficulty (Difficulty_avg) Cluster 1 represents categories where the perceived difficulty is average, suggesting that participants found the tasks neither particularly easy nor difficult. This perception reflects their standard performance. Differently, cluster 2 represents categories where participants perceived the tasks as less difficult. This perception aligns with their efficient performance, indicating that their skills and strategies made the tasks feel easier. In to the results of the other variables, cluster 3: Participants in this cluster found the tasks more difficult, consistent with their less efficient performance. The higher perceived difficulty could be due to lower skill levels, less familiarity with the tasks, or inherently challenging tasks.

Table 4 report the names of the semantic categories belonging to each cluster. Cluster 1: Leaf, Flower, LightBulb, Spider, Bee. Cluster 2: Sheep, Bus, Train, Mermaid, Ambulance, Owl, Lion, Tractor. Cluster 3: Pizza, Birthday Cake, Palm Three, Face. These groupings can help in categorizing and analyzing the items based on their complexity.

Name Label	Name	Cluster Group
1	Leaf	1
2	Flower	1
3	LightBulb	1
4	Spider	1
5	Pizza	3
6	BirthdayCake	3
7	PalmThree	3
8	Face	3
9	Bee	1
10	Sheep	2
11	Bus	2
12	Train	2
13	Mermaid	2
14	Ambulance	2
15	Owl	2
16	Lion	2
17	Tractor	2

Table 4: Names and their Cluster Groups

4 Triadic joint attention on communicating visual representation in HRI experiment

In human-robot interactions, joint attention — a shared focus among oneself, another agent, and a third element — plays a significant role. The robot's gazing behavior effectively captures partners' interest, leading to the establishment of joint attention (for a comprehensive review on robots and joint attention, see [3]). However, it remains unclear how individuals might change their sketching behavior when drawing for an observer. This experiment aims to understand how drawing activities can promote joint attention and enhance collaborative processes with robots. To explore this, we investigated for the first time whether and how individuals modify their drawing behaviour in the presence of a social (child) robot observing their task. Specifically, this experiment investigates whether joint attention phenomena occur during drawing activity with a child robot observer and how the presence of the robot affects this interaction. Furthermore, we explore the mechanisms related to "motionese" in the context of drawing to determine if and how the human drawing style changes while illustrating object categories for a robot observer.

4.1 Methods

4.1.1 Participants

53 human participants (25 M, 25 W, 3 NB) engaged with a social child-like robot in a drawing test conducted at two separate locations: Comenius University of Bratislava (UKBA, SK, N=27) and the Italian Institute of Technology (IIT, IT, N=26).

Only native speakers from both countries were recruited to ensure full comprehension of the task. Consequently, the experiment was conducted in Slovak and Italian languages. All participants provided written informed consent before taking part in the study. Ethical approval for the study was obtained from the regional ethical committee, Comitato Etico Regione Liguria, for IIT, and from the ethical committee of the Faculty of Mathematics, Physics, and Informatics for UKBA. Participants received a compensation of 10 € for their time.

4.1.2 Experimental sessions

To explore the variances in cognitive mechanisms and strategies triggered by the presence and actions of the robot, a within-subject experiment was designed. The experiment session was divided into two sessions (conditions), separated by a 5-minute break. Using their index fingers, participants were instructed to draw specific object categories on a touchscreen. They completed this task individually (individual condition) and then in the presence of the robotic agent (robot condition). *Individual Condition.* Participants were shown a picture of the robot and instructed to draw 12 different object categories (e.g., Computer, Duck, see table 5 for the complete list) with the following goal: "Make the robot in the picture understand your drawings." Three categories were requested twice to check for repetition effects. At the end of each drawing, participants rated the difficulty of the task on a Likert scale (ranging from 1 to 7).

Robot Condition. In the following condition, the robot stood in front of the participants, welcoming them by looking them in the face when they entered the room and before each drawing and looking at the screen during the drawing completion. Participants were instructed to draw 12 object categories. 6 categories out of 12 were already present in the individual condition. After participants finished drawing, the robot gave neutral vocal feedback: "Ok". This choice follows the idea that a lack of feedback could be interpreted as anti-social

behavior, while positive feedback could be an additional factor influencing participants. After each drawing, while participants scored the task's difficulty, the robot looked away to avoid exerting pressure on them. The order of conditions remained consistent throughout the initial condition and was designed to exclude the robot, thus mitigating any potential biases from encountering the robot.

4.1.3 Stimuli (object categories)

As in the previous experiment, the 18 object categories were selected from those included in the Google Quick Draw Dataset [4] [5] [60]. For this experiment, the criteria of selection for the stimuli was based on an average number of traits ranging from 6 to 10. There were no drawings with an average of fewer than 6 traits or more than 10. Table 5 displays the chosen categories based on the condition(s) in which they were presented. The stimuli presentation followed the same protocol as the stimuli validation experiment. Nonetheless, since we additionally wanted to control the repetition effect, the two conditions consisted of two different sets of stimuli. In each condition, 12 categories were presented to participants. 6 of them were proposed again in the robot condition. 3 among these 6 were proposed two times in the individual condition. This specific design was conceived to balance a) the need for category repetition (for a repeated measure comparison overcoming the problem of between-categories variability), b) the need to augment the set of stimuli proposed to participants (to enhance the generalizability of the findings to different categories), and c) the possibility to check for potential confounding effects such as the repetition of the categories over time.

Table 5: Object Categories divided for each Condition.

INDividual 15 drawings (12 categories)		**ROBot 12 drawings (12 categories)**
object categories	repeated	object categories
Bee, Bus, Sheep	Bee, Bus, Sheep	Bee, Bus, Sheep
Duck, Face Computer		Duck, Face Computer
Alarm Clock, Ambulance, Ant Crab, Drums, Penguin		Map, Mosquito, Pig, Pizza, Sea Turtle, Teddy Bear

[4] https://github.com/googlecreativelab/quickdraw-dataset

[5] https://console.cloud.google.com/storage/browser/quickdraw_dataset/full/raw;tab=objects

4.1.4 Variables and measurements

Variables to assess the task complexity are the same as the stimuli validation experiment. However, additional variables introduced in the main experiment aim to explore the effect of joint attention on the drawing task.

Latency time, Drawing time, Number of strokes do not differ from the previous experiment.

Avg. Pause. Pauses are thought of as the difference in time between the first timestamp of one stroke t_{start_i} and the last timestamp of the previous one $t_{end_{i-1}}$. It can be interpreted as the reflection time while drawing when trying to fill the gap between one's mental and graphic representation.

For what concerns the spatial features, considering each stroke formed by *m* micro-traits, the *Total Stroke Length* (*Tot.StrkL.*) computed on *n* strokes can be calculated by summing the micro-traits length of each stroke *j*, considered as the difference between the coordinates of two subsequent pixels (*i* and *i*-1):

Average Length (*Avg.Len.*) can be derived dividing the total stroke lengths for the number of strokes:

Bounding Box. The dimension of the imaginary rectangle area enclosing the final drawing, computed from the greatest and smallest pixel coordinates of the drawing The Tot.StrkL., the Avg.Len. and the BoundingBox can be used to indicate the amplitude of the drawing.

Avg. Velocity. Computed the length of the strokes and considering the contributions of the strokes' *Drawing Time*, it is immediate to find the velocity of each stroke as $\frac{length}{time}$

4.1.5 Questionnaires and scales

Participants were asked to fill out several surveys before the beginning of the experiment after observing a picture of the robot. The same questionnaires were also submitted at the end of the experiment, i.e. after the interaction.

Inclusion of Other in the Self (IOS): This single-item scale [61] measures the level of closeness the respondent experiences towards another agent or a group. The scale levels are typically 7 pairs of circles arranged from left to right. The circles on the far left have no overlap, indicating no connection or a very distant relationship. The circles on the far right have the most overlap, indicating a very close, interconnected relationship. Intermediate pairs of circles show varying degrees of overlap to represent different levels of closeness.

Respondents are asked to choose the pair of circles that best represents their relationship with the other person or entity (e.g. "Please choose the number of the image that best represents the relationship between you and the robot"). The selection reflects how much the respondent perceives the other as part of themselves (see Figure 4).

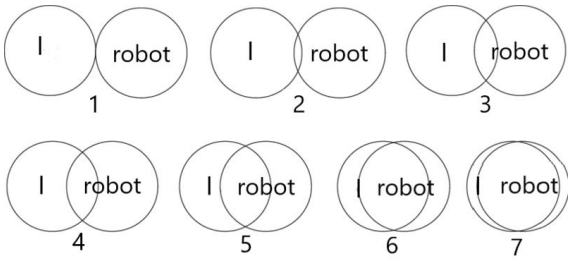

Figure 4: Visual representation of the IOS scale

Godspeed Questionnaire: The participants filled out the scales Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety [62] immediately after the IOS. The Godspeed Questionnaire was preferred to other questionnaires used in HRI because of its extensive use and availability in many languages. Since no additional HRI questionnaire was present in Slovak, it was easier to translate it with a double check from the English and the Czech versions.

Additionally, after each drawing, we asked participants to self-evaluate the difficulty they faced in drawing that specific sketch with a scroll bar ranging from 1 to 7 (*Task Difficulty Survey*). We also questioned the participants about the likability of each draw and the enjoyability of the drawing activity, as in the previous experiment. For an example of the questions see Subsection 3.1.5.

4.1.6 Setup

At both experimental sites, the rooms were partitioned into two compartments to minimize the influence of the experimenter on the participant, as illustrated in Figure 5. The experimenter could observe the experiment via two cameras installed to provide frontal and lateral views of the participant.

Within the experimental compartment, there was a chair, a desk, and an LCD Touchscreen Monitor placed atop it. The models used for the experiment were the ELO 2002L at IIT (436.9x240.7 mm, 1920x1080 px, 60 Hz) and the ELO 2202L at UKBA (476.06x267.79 mm, 1920x1080 px, 60 Hz). A cross-colored tape marked the initial hand position of the

participants before the start of the drawing session. In the robot condition, the robot was positioned in front of the participant, on the opposite side of the touchscreen.

4.1.7 The Robots

The use of humanoid robots in this study aimed to exploit the controllability and repeatability of robots' actions and their human-like, social appearance and behavior to study human cognitive mechanisms in an interactive embodied scenario. We used an iCub robot (IIT) [63] and a Nico robot (UKBA) [64] that was inspired by the iCub in its design (see Fig. 5 for a picture of the two robots).

iCub iCub is a complex robotic platform developed to study human cognition with computational and Human-Robot Interaction approaches. It has been designed with the shape of a 5-year-old human child and can engage in social interactions given its motor-cognitive and social abilities. The sensors used in this experiment are the two cameras installed in its eye cavities. The three DoFs of the neck and the three for the eyes allowed the robot to perform head-gaze movements generated with the iKinGazeCtrl module [65]. Specifically, the robot looked at the touchscreen during the drawing activity, at some random points in the room away from the participant during the question time after the drawing, and it tracked the participant's face at the beginning of the experiment and before every drawing. The LEDs to generate facial expressions were switched off to make it appear more similar to the version of Nico at UKBA. The various modules communicated through YARP middleware (v3.8)[6], [7].

Nico Nico was designed and built taking inspiration from the iCub to produce a cheaper version of it. Nico's neck two DoFs were controlled to allow the same gaze behaviors as iCub. Based on the iCub head, it has 2 overlapping cameras in its eye cavities. As for the iCub, we used the cameras for participants' Face-Tracking. Nico's version at UKBA is not endowed with legs, which, in any case, were not needed for the interaction. Thus, it was placed directly on the table. Nico's code was written in python3.86, 7.

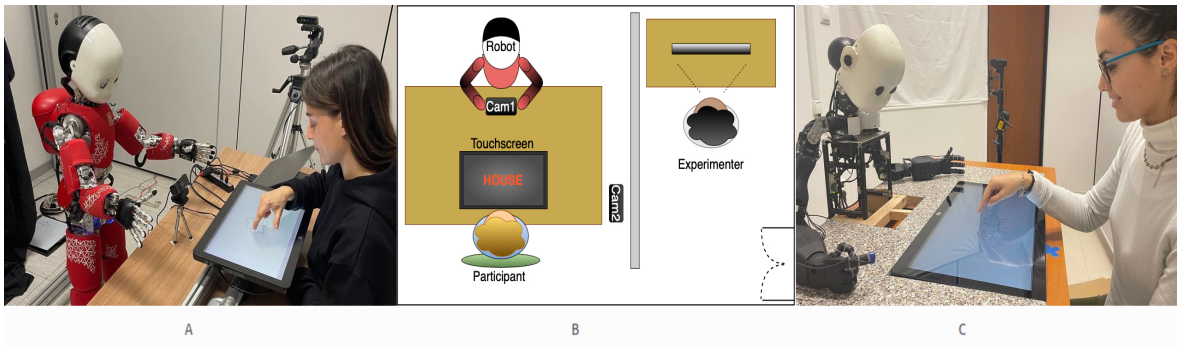

Figure 5: Pictures and Schema of the experimental setup for the robot condition: A at IIT (Italian site), B schematic representation of the setup, C at UKBA (Slovak site).

4.1.8 Data Analysis

4.1.9 Feature Extraction

To test our hypothesis, we based our analysis on quantitative measures extracted from the drawing activity (see Fig. 6). The basic components of drawings are strokes, which are drawing traits produced by a continuous touch of the finger on the screen. Strokes are defined by triplets of data (x, y, t), representing the Cartesian coordinates of the trait, x and y, measured in pixels, for each timestamp, t, measured in nanoseconds. Pixel coordinates were measured by taking the top left corner of the screen as a reference and ranging according to the screen resolution (1920x1080). Data were collected with the Python library Tkinter. Through strokes, spatial and temporal information of the drawing became available. Linking such information with the timestamp of the stimulus window used to show the object category to participants, we could extract all the parameters needed to evaluate different features of the drawing as follows.

4.2 Robot experiment results

Complexity of representing semantic categories construct – Second experiment results

4.2.1 Difficulty ranking results

Table 6 reports descriptive statistics about each category's difficulty ranking which was measured with a one-question survey after the completion of the drawing. For the words "Alarm Clock", "Ambulance", "Ant", "Bee", "Bus", "Computer", "Crab", "Drums", "Duck", "Face",

[6]DOI of codes with git links: https://doi.org/10.5281/zenodo.10944480
[7]DOI of data: https://doi.org/10.5281/zenodo.10943977

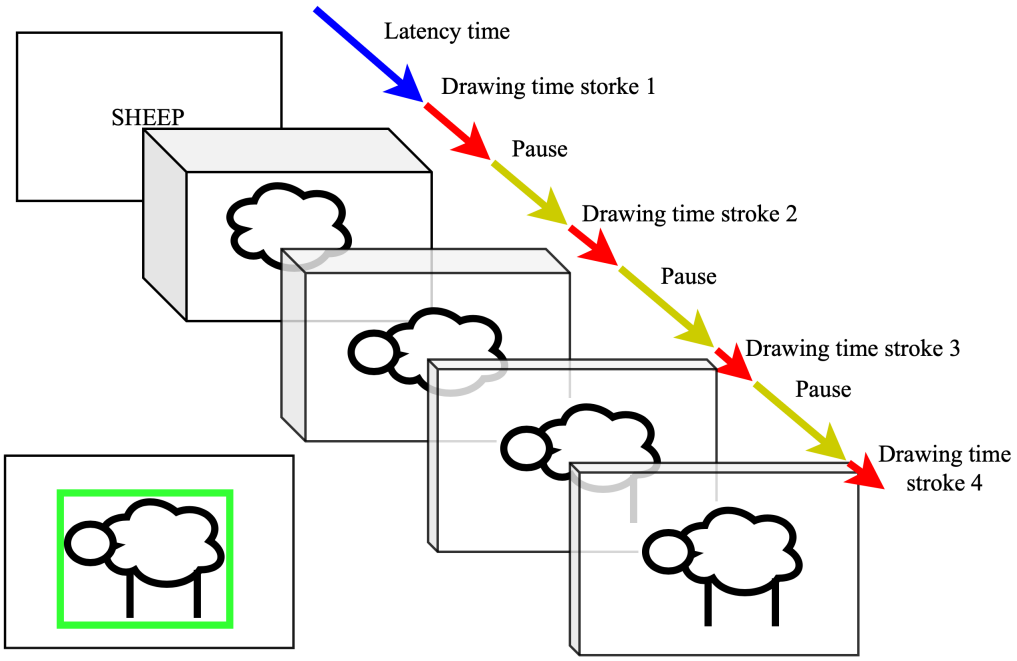

Figure 6: Graphic representation of the drawing activity with some of the extracted drawing parameters illustrated. After the window with the object category is shown to participants, Latency Time (blue arrow) is computed until the first stroke is sketched. Drawing Time (red arrows) and Pauses (yellow arrows) are calculated as shown in the picture. The green rectangle shows the bounding box for this specific drawing.

"Map", "Mosquito", "Penguin", "Pig", and "Sheep", Shapiro-Wilk tests revealed significant departures from normality at the $p < .05$ level. For "Sea Turtle," the Shapiro-Wilk test did not indicate a significant departure from normality ($p = .328$). For "Teddy Bear," while the Shapiro-Wilk test approached significance ($p = .117$), it did not reach conventional levels of significance. These results suggest that for most of the words, the difficulty rankings do not follow a normal distribution. However, for "Sea Turtle" there is no evidence of departure from normality based on the Shapiro-Wilk test.

4.2.2 Spearman correlation results

??

Table 6: Descriptive statistics of difficulty rankings

Word	**Mean**	**Median**	**Standard Deviation**	**Shapiro-Wilk p**
Alarm Clock	2.06	2.10	1.53	0.001
Ambulance	3.91	4.00	1.66	0.037
Ant	3.25	3.50	1.96	0.029
Bee	3.05	2.70	1.75	<.001
Bus	2.90	2.69	1.62	<.001
Computer	2.31	1.90	1.66	<.001
Crab	4.14	4.70	1.99	0.013
Drums	4.51	4.60	2.02	<.001
Duck	3.53	4.00	1.81	0.002
Face	2.81	2.45	1.84	<.001
Map	3.54	3.92	2.04	0.003
Mosquito	4.11	4.60	1.85	0.018
Penguin	3.80	4.20	1.95	0.020
Pig	3.29	3.40	1.71	0.004
Pizza	2.00	1.62	1.46	0.004
Sea Turtle	3.50	3.35	1.64	0.328
Sheep	3.31	3.40	1.73	<.001
Teddy Bear	3.31	3.00	1.85	0.117

Table 7: Spearman's rho correlation coefficients and p-values for the relationships between Difficulty Ranking, Latency Time, Total Time, Number of Strokes, Enjoyment Ranking, and Likeability Ranking. Significant correlations are indicated with *p < .05, **p < .01, and ***p < .001.

Spearman Correlation Matrix							
		Difficulty	Latency Time	Total Time	N Strokes	Enjoyment	Likeability
Difficulty	Spearman's rho	–					
	p-value	–					
Latency Time	Spearman's rho	0.783**	–	-			
	p-value	0.004	–				
Total Time	Spearman's rho	0.350	-0.007	–			
	p-value	0.266	0.991	–			
N Strokes	Spearman's rho	-0.210	-0.545	0.699*	–		
	p-value	0.514	0.071	0.015	–		
Enjoyment	Spearman's rho	-0.699*	-0.566	-0.028	0.196	–	
	p-value	0.015	0.059	0.939	0.543	–	
Likeability	Spearman's rho	-0.350	-0.524	0.161	0.399	0.713*	–
	p-value	0.266	0.084	0.619	0.201	0.012	–

Table 7 Spearman's rho correlation coefficient was computed to estimate the monotonic relationship between the perceived difficulty and the other variables. Table **??** shows the Spearman analysis results. Findings reported the correlations between latency time, enjoyment ranking, and perceived difficulty were statistically significant. Therefore we found a strong positive correlation between latency time and perceived difficulty, $\rho(10) = 0.783, p = 0.004$. A strong negative correlation between enjoyment ranking and perceived difficulty, $\rho(10) = -0.699, p = 0.015$. A weak negative correlation between likeability ranking and perceived difficulty, $\rho(10) = -0.350, p = 0.266$.

4.2.3 K-means Clustering analysis results

Table 8: Cluster No represents the identified clusters; Latency Time, Total Time, Number of Strokes, and Difficulty Ranking are the centroid values for each cluster.

Centroids of Clusters					
	Cluster No	Latency time	Total time	Total strokes	Difficulty
1	1.00	0.502	1.280	0.948	0.696
2	2.00	-1.187	0.008	1.110	-1.849
3	3.00	0.124	-0.551	-0.723	0.230

Table 8 reported the results of the K-means Clustering. The analysis identified three clusters with the following centroid values:

Cluster 1 had a latency time of 0.502, a total time of 1.280, a number of strokes of 0.948, and a difficulty ranking of 0.696. Cluster 2 exhibited a latency time of -1.187, a total time of 0.008, a number of strokes of 1.110, and a difficulty ranking of -1.849. Cluster 3 showed a latency time of 0.124, a total time of -0.551, a number of strokes of -0.723, and a difficulty ranking of 0.230.

Triadic joint attention on communicating visual representation in HRI experiment results

In our study, we measured various drawing features across two conditions: Individual and Robot (see Table 9. The latency time in the Individual condition (M = 3.57, SD = 3.68) was higher compared to the Robot condition (M = 2.84, SD = 1.74). The average pause was also longer in the Individual condition (M = 1.17, SD = 0.50) than in the Robot condition (M = 0.97, SD = 0.37). Drawing time was similar between the Individual (M = 17.3, SD = 8.4) and Robot conditions (M = 17.0, SD = 7.89).

Participants made slightly more strokes in the Individual condition (M = 23.5, SD = 12.5) compared to the Robot condition (M = 22.2, SD = 11.5). The total length of the drawings was

Table 9: Descriptives of drawing features in terms of averages and sd.

Drawing features	**All Object Categories**			
	INDIVIDUAL		ROBOT	
	mean	std	mean	std
Latency Time (s)	3.57	3.68	2.84	1.74
Avg. Pause (s)	1.17	0.5	0.97	0.37
Drawing Time (s)	17.3	8.4	17.0	7.89
Stroke Number (#)	23.5	12.5	22.2	11.5
Tot. Length (cm)	132.0	91.7	158.0	98.9
Avg. Length (cm)	6.61	5.01	8.13	5.08
Bounding Box (cm^2)	175.0	139.0	238.0	173.0
Avg. Velocity (cm/s)	7.97	4.01	9.78	5.0

shorter in the Individual condition (M = 132.0 cm, SD = 91.7) than in the Robot condition (M = 158.0 cm, SD = 98.9). Similarly, the average length of each stroke was shorter in the Individual condition (M = 6.61 cm, SD = 5.01) than in the Robot condition (M = 8.13 cm, SD = 5.08).

The bounding box, which measures the area occupied by the drawing, was smaller in the Individual condition (M = 175.0 cm 2 2 , SD = 139.0) compared to the Robot condition (M = 238.0 cm 2 2 , SD = 173.0). Finally, the average velocity of drawing was slower in the Individual condition (M = 7.97 cm/s, SD = 4.01) compared to the Robot condition (M = 9.78 cm/s, SD = 5.0).

4.3 Robot Experiment discussion

4.4 Complexity of representing semantic categories construct - Second experiment discussion

Figure 7 displays the difficulty ranking across the 15 categories. The results displayed in Table 6 suggest that for most of the categories, the difficulty rankings do not follow a normal distribution.

The analysis of the correlation matrix revealed several notable relationships among the measures. There was a strong positive correlation between Difficulty ranking and Latency time, suggesting that tasks perceived as more difficult took longer to initiate. Difficulty ranking also had strong negative correlations with both Enjoyment rankings, indicating that more difficult tasks were associated with lower enjoyment. Total time to complete tasks

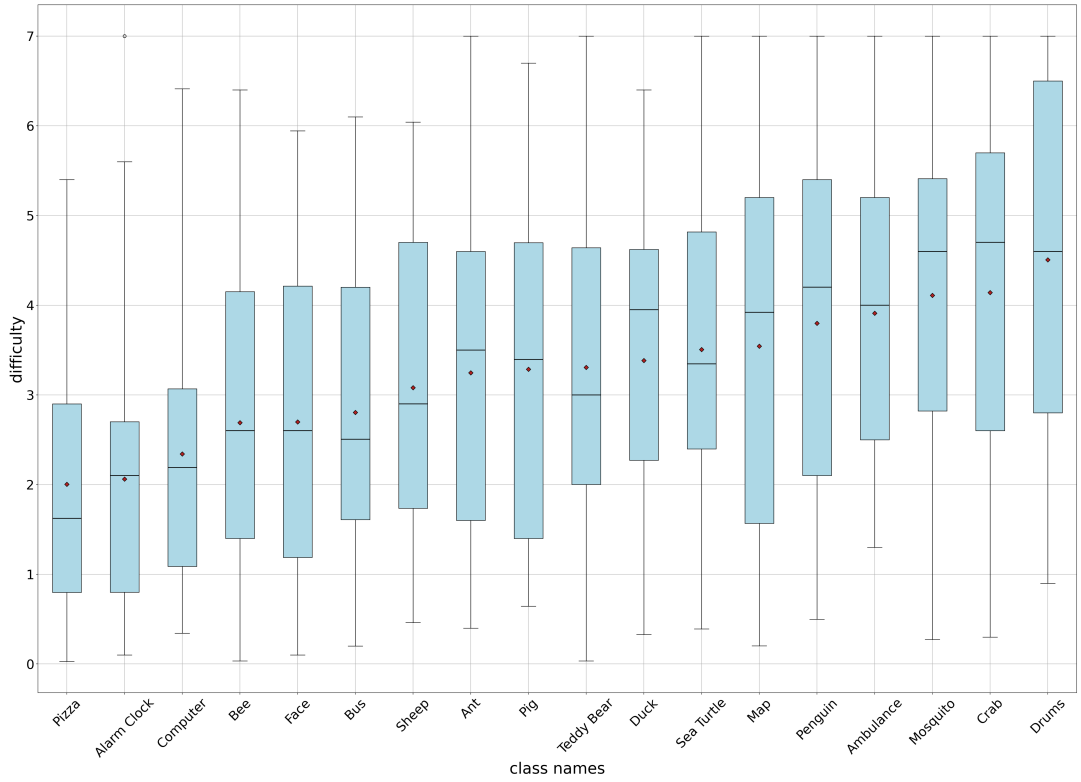

Figure 7: Comparing the reported difficulty of categories within participants (N=53)

was strongly positively correlated with the Number of Strokes, meaning that tasks taking longer also required more strokes. However, Total time did not significantly correlate with Enjoyment or Likeability rankings, indicating that the total duration of the task did not substantially affect how enjoyable or likeable the task was perceived to be. The strong positive correlation between Enjoyment and Likeability Rankings indicates that tasks that were more enjoyable were also perceived as more likeable, which is an expected outcome as enjoyment typically contributes to overall positive perceptions.

Overall, the data suggests that task difficulty plays a significant role in both the time taken to initiate tasks and participants' enjoyment levels. These findings suggested that the sensitivity related to the inherent difficulty of the single category representation influenced devising strategies to process the semantic content into its graphic representation. Additionally, the sensitivity to task difficulty influenced the perceived enjoyment related to the drawing session (the enjoyability). Regarding the correlation between the total time and number of strokes. Results indicated the mental effort is related to the number of strokes, that is representative of both the degree of visual complexity and number of actions required to complete the task. Eventually, the relationship between enjoyment and likeability under-

scores the importance of designing task that not only engage participants but also leave them with positive perceptions of the drawing outcome.

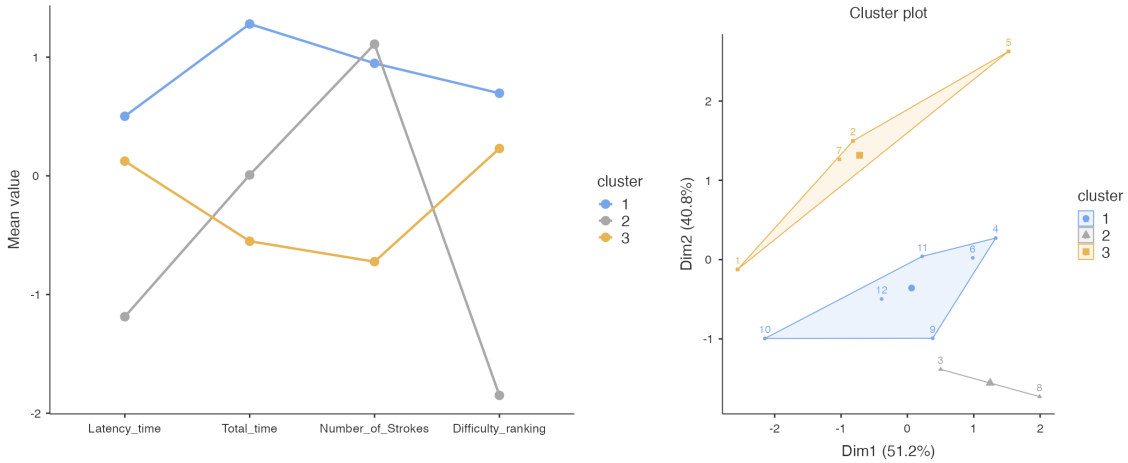

Figure 8: *Left:* graphical representation of three distinct trends based on participants' recorded latency time, total time, number of strokes, and difficulty. *Right:* plot of clustered categories with similar trends in latency time, total time, number of strokes, and difficulty.

The results summarized in Table 8 reported three clusters indicating distinct patterns in their respective drawing behaviors, as measured by latency time, total time, number of strokes, and difficulty ranking. Cluster 1 had a relatively moderate latency time (0.502), suggesting that individuals in this cluster took a reasonable amount of time to start their drawings. A relatively moderate latency indicates that for these categories participants on average were neither quick nor overly slow to devise a strategy and begin. The total time is quite high = 1.280, indicating that participants in Cluster 1 spent a considerable amount of time on their drawings. This might suggest a careful, detailed approach to the task due to major mental effort investment. A high number of strokes (0.948) implies that drawings in this cluster were relatively detailed and complex. The difficulty ranking is moderate (0.696) suggesting that for these categories participants found the drawing task somewhat challenging but manageable, in terms of attentive resources demand.

Cluster 2 had a significantly negative latency time (-1.187), meaning these participants were very quick to start their drawings, possibly indicating a high level of skills strategy. The average total time is almost zero (0.008) indicating that drawings categories belonging to this cluster were completed very quickly. This could suggest a low amount of attentive resources required to complete these sketches. Despite the quick completion, the high number of strokes (1.110) suggests that their drawings were quite detailed. This combination of low speed and high detail might indicate these categories are on average easier to repre-

sent compared to the other clusters, likely due to familiarity of these semantic categories. Accordingly, the significantly negative difficulty ranking (-1.849) indicates that participants average found the categories in this cluster very easy to draw. Therefore the sensitivity to task difficulty was low.

Cluster 3 exhibited a slightly positive latency time (0.124), suggesting these participants were moderately quick to start their drawings. A negative total time (-0.551) suggests that participants completed their drawings relatively quickly, though not as fast as those in Cluster 2. The low number of strokes (-0.723) indicates that their drawings were less detailed and more simplistic compared to the other clusters. The moderate difficulty ranking (0.230) suggests that participants in this cluster found the task neither particularly easy nor difficult.

The Figure 8 accurately reflects the results reported in the table 8. The plot on the left displays the mean values of four variables – Latency time, Total time, Number of Strokes, and Difficulty ranking – across three different clusters. Cluster 1, 2 and 3 are respectively represented with blue, grey and orange lines. Overall Trends

Latency time: Cluster 1 exhibits the highest latency time, with a positive mean value indicating longer delays before task initiation compared to other clusters. Cluster 3 shows a moderate latency time with a mean value around zero, indicating average task initiation times. Cluster 2 exhibits the lowest, with a significantly negative mean value, indicating quick task initiation. This likely because devising a strategy to represent these categories was easier compared the other clusters.

Total time: Cluster 1 has the highest total time for task completion, suggesting that categories in this cluster took the longest amount of time and attentive resources to be completed. Whereas total time in Cluster 3 is negative and the lowest. This suggest that categories in this cluster are completed faster than in the other clusters. For Cluster 2 the Total time is close to zero, suggesting that the categories take an average amount of time and attentive resources to complete.

Number of Strokes: Cluster 2 requires the highest number of strokes, indicating more complex visual representation. Oppositely, the categories in Cluster 3 are characterized by the fewest Number of Strokes – that is also negative – suggesting the visual representation of these categories was less complex compare the other clusters. The Number of Strokes for Cluster 1 is moderate, with a mean value slightly above zero, indicating these categories were represented with a relatively balanced amount of interacting number of details, revealing

moderate task complexity.

Difficulty ranking: Cluster 2 perceives the tasks as the least difficult, whereas Cluster 3 finds them the most difficult. Therefore, the Difficulty ranking for Cluster 2 was the lowest, with a significantly negative mean value. This suggests that for these categories participants were less sensitive to the intrinsic difficulty of representing these semantic categories. for Cluster 3 the Difficulty ranking is positive, indicating that tasks are perceived as more difficult compared to the other clusters.

The cluster plot on the right shows the distribution of data points across three clusters in a two-dimensional space defined by two principal components (Dim1 and Dim2). Each data point within the clusters is labeled for identification. Data points in Cluster 1 are labeled 4, 6, 9, 10, 11, and 12. Data points in Cluster 2 are labeled 3 and 8. Data points in Cluster 3 are labeled 1, 2, 5, and 7. Cluster 1 (Blue): Contains more data points and covers a larger area in the lower-left quadrant, extending slightly into the upper right, indicating more variability within this cluster. Cluster 2 (Grey): Contains fewer data points, closely grouped in the lower-right quadrant, suggesting this cluster is more compact and homogeneous. Cluster 3 (Orange): Also contains fewer data points but is spread out in the upper-left quadrant, indicating some variability within this cluster.

Table 10: Name categories and corresponding clusters from K-means Cluster analysis

Category label	**Category name**	**Cluster**
1	Alarm Clock	2
2	Ambulance	3
3	Ant	1
4	Crab	1
5	Drums	3
6	Penguin	1
7	Map	3
8	Mosquito	1
9	Pig	1
10	Pizza	2
11	Sea Turtle	1
12	Teddy Bear	1

Table 10 reports the categories per each cluster. Notably, the three clusters are associated to three conceptual classification. Cluster 1: "Animals" - This cluster includes categories such as "Ant," "Crab," "Penguin," "Mosquito," "Pig," "Sea Turtle," and "Teddy Bear," indicating that these categories are likely related to various animals. Cluster 2: "Objects" -

This cluster includes categories such as "Alarm Clock" and "Pizza," suggesting that these categories are associated with inanimate objects. Cluster 3: "Events/Actions" - This cluster includes categories such as "Ambulance," "Drums," and "Map," implying that these categories are linked to events or actions rather than specific objects or animals

4.5 Triadic joint attention on communicating visual representation in HRI experiment discussion

To evaluate the impact of the robot's presence and behavior on participants' drawing activity and graphic representations, we conducted a quantitative analysis by comparing the two primary conditions of our experiment: individual versus robot. We assessed several parameters to gauge the differences between these conditions.

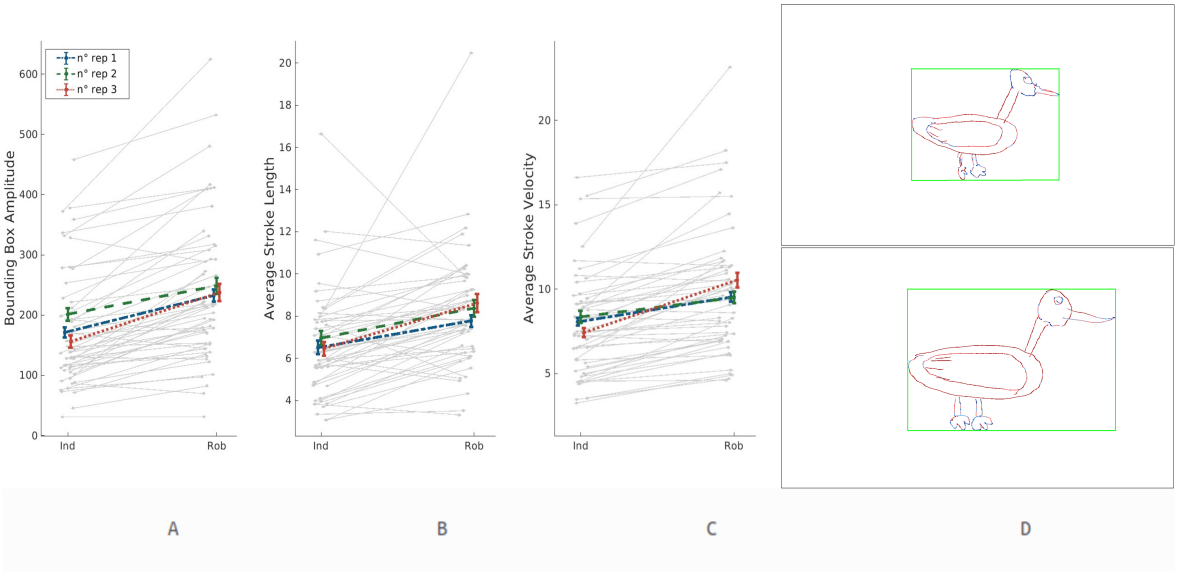

Figure 9: Plots representing the mean and standard error (error bars) values for Bounding Box (A), Avg. Length (B) and Avg. Velocity (C) for every participant, both in INDividual and ROBot conditions. Global means of these features are shown for the object categories represented 1, 2, and 3 times. Grey dots and lines show each participant's means for the two conditions. Figure D shows the two drawings for 'Duck' of one representative participant, one sketched in the Individual (upper one), the other in the Robot condition (bottom one). The two drawings show the larger Bounding Box, longer Avg. Length, and faster Avg. Velocity for the robot condition. The two Ducks are colored using two colors, two shades per color: red for faster micro-traits (the darker, the faster), and blue for slower micro-traits (the darker, the slower).

We performed a statistical analysis to evaluate the Condition effect, representing the influence of the robot's presence and joint attention behavior. We considered this effect a predictor for each feature (dependent variables) described in Section 4.1.8. This was ac-

complished by fitting eight distinct Linear Mixed Models (LMM), with each model corresponding to one of the drawing parameters outlined in Section 4.1.8 (averages and standard deviations in Table 9). This approach enabled us to utilize the same model for assessing the effects of additional potential predictors, such as the Number of Representations (i.e., the frequency of representing an object category by a participant, indicating a repetition effect), and the Experimental Site. Furthermore, we incorporated random effects for other factors, including participants and object categories.

The random effect of participants was included to account for individual baseline differences and to model the intra-subject correlation arising from repeated measurements. Similarly, the random effect at the object category level was introduced to capture inter-stimulus variability in error parameters. These random effects were included in the model in the following order.

We computed the shift of the robot condition from the individual one, considered as a baseline. We found a significant decrease in the Latency Time (ROB – IND: B=-0.801, t=-4.188, p<0.001) and the Avg. Pause (ROB – IND: B=-0.209, t=-7.197, p<0.001) and a significant increase in the Tot. Length (ROB – IND: B=27.760, t=3.448, p=0.002), the Avg. Length (Ind – Rob: B=1.61, t=4.563, p<0.001), the Avg. Velocity (ROB – IND: B=1.921, t=6.508, p<0.001), and the Bounding Box (ROB – IND: B=63.25, t=4.93, p<0.001) (Fig. 9 graphically shows the Tot. Length, Avg. Velocity, and Bounding Box results). After applying the Bonferroni correction to our tests, post hoc results confirmed the significance of the effects. No significant effect of Condition was found for the Stroke Number and the Drawing Time.

Furthermore, we analyzed the questionnaires administered to participants both before the experimental phase and after completing all tasks. These measures aimed to investigate whether differences in participants' drawing representations between conditions could be attributed to their perceptions of the robot. To achieve this, we examined participants' evaluations of the robot, as measured by the IOS and Godspeed scales. We assessed differences between participants' pre-experiment expectations and post-experiment ratings and tested whether such differences correlated with variations in participants' drawing features from the individual to the robot condition.

The only statistically significant difference identified by paired T-Tests (or Wilcoxon signed-rank tests when necessitated by non-normal distributions) was observed in God-

speed's Anthropomorphism scale. Specifically, there was a decrease from the pre-experiment (M=14.5, Std=3.69) to the post-experiment rating (M=11.2, Std=4.21). No other significant differences were found in the tests conducted before and after the interaction with the robot for all other scales (all ps > 0.05).

In the correlation analysis, Spearman's rank correlation unveiled a significant positive correlation between delta-IOS (the difference between post-experiment and pre-experiment) and delta-Bounding Box (the difference between the robot and individual conditions): r(51)=0.361, p=0.045, as depicted in Fig. 10. Similarly, a positive correlation was observed between delta-IOS and Delta-Drawing Time (r(51)=0.320, p=0.020). Delta-Bounding Box and Delta-Drawing Time were calculated by averaging all categories for each participant. This trend persisted throughout the analysis.

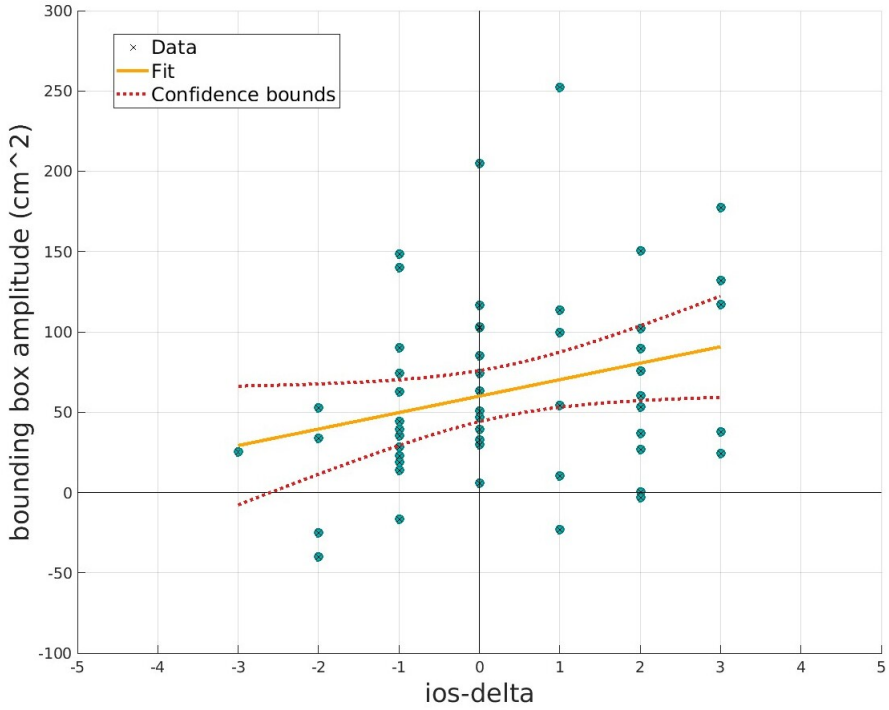

Figure 10: The scatter plot illustrates the relationship between the shift in Bounding Box (Delta-Bounding Box, computed as robot-individual) and the shift in IOS (Delta-IOS, computed as a post-pre experiment). Each data point represents a participant's response. A yellow linear fit line is superimposed on the data to highlight the positive linear correlation between the two variables. This plot demonstrates how changes in participants' perception of the robot (measured by Delta-IOS) correspond to changes in the size of the bounding box of their drawings (measured by Delta-Bounding Box). The positive linear fit indicates that as participants' perception of the robot increases, there tends to be a corresponding increase in the size of the bounding box of their drawings.

4.6 Repetition effect check

To assess the impact of repetition concerning the robot effect, we conducted a detailed examination of the results obtained from the eight aforementioned Linear Mixed Models (LMMs). Upon analysis, no statistical significance was observed regarding the effect of the Number of Representations and its interaction with the Condition, except for three parameters: Avg. Pause, Tot. Length, and Avg. Velocity.

Specifically, concerning Avg. Pause, a significant interaction effect between the Number of Representations and the Condition was evident for categories drawn three times compared to those drawn two or one time. In the robot condition, there was a notably greater decrease in Avg. Pause for categories represented three times (ROB – IND * 2 – 3: B=0.166, t=2.855, p=0.004, and ROB – IND * 1 – 3: B=0.168, t=2.182, p=0.037).

As for Tot. Length and Avg. Velocity, the interaction effect exhibited an opposite trend between categories drawn three and two times, with a significantly greater increase in these parameters in the robot condition for categories represented three times (Tot. Length: ROB – IND * 2 – 3: B=-32.159, t=-2.862, p=0.004, Avg. Velocity: ROB – IND * 2 – 3: B=-1.966, t=-4.199, p<0.001).

For a comprehensive analysis, we also examined whether similar effects persisted when considering only categories represented once (where no repetition/habit effect is present). Notably, we found significant differences in both features (Avg. Velocity: ROB - IND: t=5.575, p<0.001, Tot. Length: ROB - IND: t=3.463, p<0.001). This consistency underscores the robustness of our findings.

We examined whether different experimental sites and types of robots influenced the drawing parameters and the robot effects discussed earlier. The only drawing feature that exhibited variability across different experimental sites was the Avg. Length, which was notably higher for drawings completed in Slovak (SK) compared to Italian (IT) conditions (SK – IT: B=1.426, t=2.372, p<0.021). However, no interaction effect of the Experimental Site with the Condition (individual/robot) was observed.

Subsequently, to assess any relationship between participants' self-reported Drawing Difficulty and the aforementioned drawing features, we conducted a Spearman's rank correlation by averaging the results for each category. A positive correlation was identified between self-reported Drawing Difficulty and Latency Time (r(10)=0.783, p=0.004), as well as with Avg. Pause (r(10)=0.708, p=0.010), but not with other drawing features. Additionally, a pos-

itive correlation was found between Latency Time and Avg. Pause: r(10)=0.620, p=0.032. These findings provide insight into participants' perceptions of drawing difficulty and its association with specific drawing parameters.

5 General discussion

Experiment 1. The first experiment in this study aimed to evaluate the complexity of representing semantic object categories by assessing cognitive load indicators. The experiment first focused on the correlation between perceived difficulty and various performance metrics, afterwards they explored clusters of drawing behaviors and their relationship to cognitive load. The findings from the two data collection highlighted several consistent patterns. First, there is a significant positive correlation between perceived difficulty and latency time, total time, and the number of strokes. This suggests that tasks perceived as more difficult require more time to initiate, more time to complete, and involve more detailed and complex drawings. Conversely, enjoyment and likability exhibited negative correlations with perceived difficulty, indicating that participants enjoyed and liked the task outcomes less when they found the tasks more difficult. The clustering analyses further supported these findings by identifying three distinct clusters of drawing behaviors. Cluster 1 was characterized by higher latency times, total times, and a moderate number of strokes, suggesting that these tasks were perceived as moderately challenging and required significant cognitive resources. Cluster 2 had the lowest latency and total times but the highest number of strokes, indicating efficient but detailed task execution. Cluster 3 showed moderate latency, lower total times, and fewer strokes, indicating simpler and less challenging tasks.

The hypotheses were confirmed as the chosen cognitive load indicators—perceived difficulty, latency time, total time, and the number of strokes—were effective in assessing the complexity of representing semantic object categories. The strong correlations between these indicators and the clustering analysis provide robust evidence that these measures can reliably assess task complexity and cognitive load. The findings align with previous research on Cognitive Load Theory (CLT), which emphasizes the relationship between task complexity, cognitive load, and performance metrics. The observed correlations between task difficulty, time on task, and mental effort are consistent with the theoretical framework of CLT, which posits that more complex tasks impose higher intrinsic cognitive loads and require more cognitive resources [31, 32]. Moreover, the study's use of subjective ratings and behavioral data to assess cognitive load reflects the methodologies recommended by recent CLT research, which advocates for a multi-faceted approach to measuring cognitive load [33].

Experiment 2. The present study investigated the impact of joint attention on social cognition and human-robot interaction, specifically focusing on how the presence of a child-like robot observer influences individuals' drawing behaviors. The study aimed to explore the potential modifications in drawing style due to the observer's presence, examining whether similar phenomena to motionese occur during this activity. *Eye Gazing and Joint Attention.* Our first hypothesis posited that the eye gazing behavior of the child-robot would effectively gain joint attention and affect participants' perception of closeness to the robot. This hypothesis was confirmed, as the data revealed significant behavioral changes in the presence of the robot, such as a decrease in latency time and average pause, along with an increase in total length, average length, average velocity, and bounding box size. These changes indicate that participants adjusted their drawing behaviors when the robot was observing, reflecting a successful establishment of joint attention. *Influence on Communication Strategies:* The second hypothesis suggested that joint attention would influence communication strategies during the drawing task for a child-robot observer. This hypothesis was also confirmed, as evidenced by the significant modifications in drawing parameters, suggesting that participants altered their communication approach when interacting with the robot. Our findings align with previous research indicating the critical role of joint attention in social cognition and interaction. For example, studies have shown that joint attention supports word learning [50], cooperation [50], and predicts social abilities [49]. The observed behavioral modifications in our study, such as changes in drawing style and increased engagement, are consistent with the established notion that joint attention facilitates communication and interaction, not only among humans but also in human-robot interactions [48].

The role of gaze in establishing joint attention [49] and the importance of gaze monitoring for predicting joint action [50] were corroborated by our results, highlighting the effectiveness of the robot's gaze behavior in engaging participants. Furthermore, the concept of motionese, where humans modify their behavior to facilitate learning in robots [57], was reflected in the drawing modifications observed in this study.

5.1 Limitations

Despite the study's strengths, several limitations should be acknowledged.

Experiment 1. First, the study relied primarily on subjective ratings and performance metrics, which can be prone to biases and may not capture the full complexity of cognitive

load. Integrating additional measures, such as physiological indicators like pupil dilation and eye tracking, could provide a more comprehensive assessment of cognitive load. These measures can offer real-time, objective data on attentional and cognitive processes, thereby complementing subjective and behavioral metrics. Moreover, the study did not consider control on familiarity with specific semantic category as confounding variable nor cross-cultural comparison in methods.

Experiment 2. The study did not fully account for individual differences in participants' familiarity and comfort with technology and robots. These differences could have impacted their interaction style and the extent to which they engaged in joint attention with the robot observer. By acknowledging these limitations and employing strategies such as the Negative Attitude Towards Robots (NARS) questionnaire to assess and control for biases, future research can enhance the robustness and validity of findings related to joint attention and human-robot interaction.

6 Conclusion

This thesis has explored the significance of joint attention in social cognition and human-robot interaction, with a particular focus on the impact of a child-like robot observer on human drawing behaviors. The research was divided into two primary studies, each contributing valuable insights to the broader understanding of cognitive load and joint attention. The first study confirmed that the selected cognitive load indicators are effective in assessing the complexity of representing semantic object categories. These findings align with previous Cognitive Load Theory (CLT) research, highlighting the importance of a multi-faceted approach to measuring cognitive load. This study provides a solid foundation for future research to further explore and refine methods for assessing cognitive load in complex tasks, despite certain limitations such as sample size and task specificity. The second study focused on the role of joint attention in social cognition and its impact on human-robot interaction. The presence of a child-like robot observer was found to significantly influence participants' drawing behaviors, confirming that joint attention can be effectively established between humans and robots. These results have important implications for the design of social robots, particularly in educational and therapeutic contexts, where enhancing collaborative processes and communication strategies is crucial. The combined findings from both studies underscore the importance of joint attention and cognitive load measurement in understanding and improving human-robot interactions. The research highlights the need for a nuanced approach to designing social robots that can effectively engage in joint attention and support complex cognitive tasks. In conclusion, this thesis has laid a robust groundwork for future research in cognitive load assessment and joint attention in human-robot interactions, emphasizing the potential of social robots to enhance collaborative and educational processes.

List of Figures

List of Tables

References

[1] J. E. Fan, W. A. Bainbridge, R. Chamberlain, and J. D. Wammes, "Drawing as a versatile cognitive tool," *Nature Reviews Psychology*, vol. 2, pp. 556–568, July 2023.

[2] A. Philippsen, S. Tsuji, and Y. Nagai, "Quantifying developmental and individual differences in spontaneous drawing completion among children," *Frontiers in Psychology*, vol. 13, p. 783446, 2022.

[3] P. Chevalier, K. Kompatsiari, F. Ciardo, and A. Wykowska, "Examining joint attention with the use of humanoid robots-a new approach to study fundamental mechanisms of social cognition," *Psychonomic Bulletin amp; Review*, vol. 27, pp. 217–236, Dec. 2019.

[4] A. Sciutti and G. Sandini, "Interacting with robots to investigate the bases of social interaction," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 12, pp. 2295–2304, 2017.

[5] *Blackwell Handbook of Social Psychology: Intraindividual Processes*. Wiley, Jan. 2001.

[6] D. Gilbert, S. Fiske, and G. Lindzey, *The Handbook of Social Psychology*. No. v. 1 in The Handbook of Social Psychology, McGraw-Hill, 1998.

[7] L. Fernandino, J.-Q. Tong, L. L. Conant, C. J. Humphries, and J. R. Binder, "Decoding the information structure underlying the neural representation of concepts," *Proceedings of the National Academy of Sciences*, vol. 119, no. 6, p. e2108091119, 2022.

[8] J. Wyer, "Principles of mental representation," *Social Psychology: Handbook of Basic Principles*, pp. 285–307, 01 2007.

[9] *The Wiley Handbook on the Cognitive Neuroscience of Learning*. Wiley, June 2016.

[10] C. Luzzatti, I. Mauri, S. Castiglioni, M. Zuffi, C. Spartà, F. Somalvico, and M. Franceschi, "Evaluating semantic knowledge through a semantic association task in individuals with dementia," *American Journal of Alzheimer's Disease & Other Dementias®*, vol. 35, p. 1533317520917294, 2020. PMID: 32308008.

[11] Y. Miyashita and T. Hayashi, "Neural representation of visual objects: encoding and top-down activation," *Current Opinion in Neurobiology*, vol. 10, no. 2, pp. 187–194, 2000.

[12] A. Martin, "The representation of object concepts in the brain," *Annual Review of Psychology*, vol. 58, p. 25–45, Jan. 2007.

[13] Y. Shtyrov, A. Efremov, A. Kuptsova, T. Wennekers, B. Gutkin, and M. Garagnani, "Breakdown of category-specific word representations in a brain-constrained neurocomputational model of semantic dementia," *Scientific Reports*, vol. 13, Nov. 2023.

[14] J. Snyder, "Drawing practices in image-enabled collaboration," pp. 741–752, 02 2013.

[15] A. Taylor, *Foreword - Re: Positioning Drawing*, pp. 9–12. Intellect Books, 2008.

[16] E. Sober, "Mental representations," *Synthese*, vol. 33, no. 1, pp. 101–148, 1976.

[17] R. Gal, Y. Vinker, Y. Alaluf, A. H. Bermano, D. Cohen-Or, A. Shamir, and G. Chechik, "Breathing life into sketches using text-to-video priors," 2023.

[18] C. Elsen, F. Darses, and P. Leclercq, *What Do Strokes Teach Us about Collaborative Design?*, p. 114–125. Springer Berlin Heidelberg, 2012.

[19] J. E. Fan, D. L. K. Yamins, and N. B. Turk-Browne, "Common object representations for visual production and recognition," *Cognitive Science*, vol. 42, p. 2670–2698, Aug. 2018.

[20] J. G. Snodgrass and M. Vanderwart, "A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity.," *Journal of Experimental Psychology: Human Learning and Memory*, vol. 6, no. 2, p. 174–215, 1980.

[21] J. Yang and J. Fan, "Visual communication of object concepts at different levels of abstraction," *Journal of Vision*, vol. 21, p. 2951, Sept. 2021.

[22] E. Rosch, C. Mervis, W. Gray, D. Johnson, and P. Braem, "Basic objects in natural categories," *Cognitive Psychology - COG PSYCHOL*, vol. 8, pp. 382–439, 07 1976.

[23] R. S. Nickerson, "Short-term memory for complex meaningful visual configurations: A demonstration of capacity.," *Canadian Journal of Psychology / Revue canadienne de psychologie*, vol. 19, no. 2, p. 155–160, 1965.

[24] R. D. Hawkins, M. Sano, N. D. Goodman, and J. E. Yang, "Visual resemblance and interaction history jointly constrain pictorial meaning," *Nature Communications*, vol. 14, Apr. 2023.

[25] T. de Jong, "Cognitive load theory, educational research, and instructional design: some food for thought," *Instructional Science*, vol. 38, p. 105–134, Aug. 2009.

[26] K. E. DeLeeuw and R. E. Mayer, "A comparison of three measures of cognitive load: Evidence for separable measures of intrinsic, extraneous, and germane load.," *Journal of Educational Psychology*, vol. 100, p. 223–234, Feb. 2008.

[27] J. Sweller, "Cognitive load theory, learning difficulty, and instructional design," *Learning and Instruction*, vol. 4, p. 295–312, Jan. 1994.

[28] E. W. Anderson, K. C. Potter, L. E. Matzen, J. F. Shepherd, G. A. Preston, and C. T. Silva, "A user study of visualization effectiveness using eeg and cognitive load," *Computer Graphics Forum*, vol. 30, no. 3, pp. 791–800, 2011.

[29] O. Chen, J. C. Castro-Alonso, F. Paas, and J. Sweller, "Undesirable difficulty effects in the learning of high-element interactivity materials," *Frontiers in Psychology*, vol. 9, Aug. 2018.

[30] S. Kalyuga, "Cognitive load theory: How many types of load does it really need?," *Educational Psychology Review*, vol. 23, p. 1–19, Jan. 2011.

[31] J. Sweller, *Cognitive Load Theory*, p. 37–76. Elsevier, 2011.

[32] F. G. W. C. Paas, J. J. G. van Merriënboer, and J. J. Adam, "Measurement of cognitive load in instructional research," *Perceptual and Motor Skills*, vol. 79, p. 419–430, Aug. 1994.

[33] E. Hoch, Y. Sidi, R. Ackerman, V. Hoogerheide, and K. Scheiter, "Comparing mental effort, difficulty, and confidence appraisals in problem-solving: A metacognitive perspective," *Educational Psychology Review*, vol. 35, May 2023.

[34] R. Brünken, T. Seufert, and F. Paas, *Measuring Cognitive Load*, p. 181–202. Cambridge University Press, Apr. 2010.

[35] O. Chen, F. Paas, and J. Sweller, "A cognitive load theory approach to defining and measuring task complexity through element interactivity," *Educational Psychology Review*, vol. 35, June 2023.

[36] P. Antonenko, F. Paas, R. Grabner, and T. van Gog, "Using electroencephalography to measure cognitive load," *Educational Psychology Review*, vol. 22, p. 425–438, Apr. 2010.

[37] R. E. Wood, "Task complexity: Definition of the construct," *Organizational Behavior and Human Decision Processes*, vol. 37, no. 1, pp. 60–82, 1986.

[38] R. Ackerman and V. Thompson, *Meta-Reasoning: What Can We Learn from Meta-Memory?* 01 2014.

[39] A. Koriat, H. Ma'ayan, and R. Nussinson Levy-Sadot, "The intricate relationships between monitoring and control in metacognition: Lessons for the cause-and-effect relation between subjective experience and behavior," *Journal of experimental psychology. General*, vol. 135, pp. 36–69, 02 2006.

[40] P. Barrouillet, S. Bernardin, S. Portrat, E. Vergauwe, and V. Camos, "Time and cognitive load in working memory," *Journal of experimental psychology. Learning, memory, and cognition*, vol. 33, pp. 570–85, 05 2007.

[41] J. Lee, "Time-on-task as a measure of cognitive load in tblt," *The Journal of AsiaTEFL*, vol. 16, p. 958–969, Sept. 2019.

[42] A. D. Bender, H. L. Filmer, K. G. Garner, C. K. Naughtin, and P. E. Dux, "On the relationship between response selection and response inhibition: An individual differences approach," *Attention, Perception, amp; Psychophysics*, vol. 78, p. 2420–2432, July 2016.

[43] F. Echenique and K. Saito, "Response time and utility," *Journal of Economic Behavior Organization*, vol. 139, pp. 49–59, 2017.

[44] J. T. Townsend and A. Eidels, "Workload capacity spaces: A unified methodology for response time measures of efficiency as workload is varied," *Psychonomic Bulletin amp; Review*, vol. 18, p. 659–681, May 2011.

[45] R. J. Innes, N. J. Evans, Z. L. Howard, A. Eidels, and S. D. Brown, "A broader application of the detection response task to cognitive tasks and online environments," *Human Factors*, vol. 63, no. 5, pp. 896–909, 2021. PMID: 32749155.

[46] K. Ouwehand, A. v. d. Kroef, J. Wong, and F. Paas, "Measuring cognitive load: Are there more valid alternatives to likert rating scales?," *Frontiers in Education*, vol. 6, 2021.

[47] N. Akhtar and M. A. Gernsbacher, "Joint attention and vocabulary development: A critical look," *Language and Linguistics Compass*, vol. 1, p. 195–207, Apr. 2007.

[48] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, "Understanding and sharing intentions: The origins of cultural cognition," *Behavioral and Brain Sciences*, vol. 28, p. 675–691, Oct. 2005.

[49] P. Mundy and L. Newell, "Attention, joint attention, and social cognition," *Current Directions in Psychological Science*, vol. 16, p. 269–274, Oct. 2007.

[50] C. O'Madagain and M. Tomasello, "Joint attention to mental content and the social origin of reasoning," *Synthese*, vol. 198, p. 4057–4078, Aug. 2019.

[51] C. Vesper, E. Abramova, J. Bütepage, F. Ciardo, B. Crossey, A. Effenberg, D. Hristova, A. Karlinsky, L. McEllin, S. R. R. Nijssen, L. Schmitz, and B. Wahn, "Joint action: Mental representations, shared information and general mechanisms for coordinating with others," *Frontiers in Psychology*, vol. 07, Jan. 2017.

[52] H. F. Chame, A. Clodic, and R. Alami, "Top-jam: A bio-inspired topology-based model of joint attention for human-robot interaction," in *2023 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, May 2023.

[53] N. SEBANZ, H. BEKKERING, and G. KNOBLICH, "Joint action: bodies and minds moving together," *Trends in Cognitive Sciences*, vol. 10, pp. 70–76, Feb. 2006.

[54] A. Clodic, E. Pacherie, R. Alami, and R. Chatila, *Key Elements for Human-Robot Joint Action*, pp. 159–177. May 2017.

[55] R. J. Brand, D. A. Baldwin, and L. A. Ashburn, "Evidence for 'motionese': modifications in mothers' inyang 2021t-directed action," *Developmental Science*, vol. 5, pp. 72–83, Mar. 2002.

[56] K. Pitsch, A.-L. Vollmer, and M. Mühlig, "Robot feedback shapes the tutor's presentation: How a robot's online gaze strategies lead to micro-adaptation of the human's conduct," *Asymmetry and adaptation in social interaction*, vol. 14, pp. 268–296, July 2013.

[57] Y. Nagai, A. Nakatani, and M. Asada, "How a robot's attention shapes the way people teach," in *in Proc. 10th Int. Conf. Epigenetic Robot*, pp. 81–88, 2010.

[58] Y. Nagai and K. Rohlfing, "Computational analysis of motionese toward scaffolding robot action learning," *IEEE Transactions on Autonomous Mental Development*, vol. 1, pp. 44–54, May 2009.

[59] A. L. Vollmer, K. S. Lohan, K. Fischer, Y. Nagai, K. Pitsch, J. Fritsch, K. J. Rohlfing, and B. Wredek, "People modify their tutoring behavior in robot-directed interaction for action learning," in *IEEE 8th International Conference on Development and Learning*, pp. 1–6, 2009.

[60] J. Jongejan, H. Rowley, T. Kawashima, J. Kim, and N. Fox-Gieg, "The quick, draw!-ai experiment," *Mount View, CA, accessed Feb*, vol. 17, no. 2018, p. 4, 2016.

[61] A. Aron, E. N. Aron, and D. Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness.," *Journal of Personality and Social Psychology*, vol. 63, pp. 596–612, Oct. 1992.

[62] C. Bartneck, *Godspeed Questionnaire Series: Translations and Usage*, pp. 1–35. Springer International Publishing, 2023.

[63] G. Metta, L. Natale, F. Nori, G. Sandini, D. Vernon, L. Fadiga, C. Von Hofsten, K. Rosander, M. Lopes, J. Santos-Victor, *et al.*, "The icub humanoid robot: An open-systems platform for research in cognitive development," *Neural networks*, vol. 23, no. 8-9, pp. 1125–1134, 2010.

[64] M. Kerzel, E. Strahl, S. Magg, N. Navarro-Guerrero, S. . Heinrich, and S. Wermter, "Nico—neuro-inspired companion: A developmental humanoid robot platform for multimodal interaction," in *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp. 113–120, IEEE, 2017.

[65] A. Roncone, U. Pattacini, G. Metta, and L. Natale, “A cartesian 6-dof gaze controller for humanoid robots.,” in *Robotics: science and systems*, vol. 2016, 2016.