Department of Education
University of Oxford

*Alexa, what are you? Exploring Primary School Children's Ontological
Perceptions of Digital Voice Assistants in Open Interactions*

Janik Festerling^a & Iram Siraj^b

^a *Corresponding author:* Janik.festerling@education.ox.ac.uk, Department of Education, University of Oxford, 15 Norham Gardens, Oxford, OX2 6PY, UK

^b Iram.siraj@education.ox.ac.uk, Department of Education, University of Oxford, 15 Norham Gardens, Oxford, OX2 6PY, UK

Abstract

Today's children grow up in an environment that is increasingly characterised by Digital Voice Assistant's (DVAs), such as Alexa, Siri or the Google Assistant. This paper argues that any attempt to investigate children's interactions with, and perceptions of, DVAs should be based on the theoretical grounds of an ontological framework that considers children's genuine understanding of what it means to be human and what it means to be a machine. Based on focus groups and a gamified data collection design, our empirical inquiry applied qualitative methods to explore primary school children's ($n=27$, age range: 6 to 10 years, average age: 8.6 years) open interactions with DVAs. In particular, our focus was how DVAs were embedded in children's general ontological belief system, and how children interpreted certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid. On the one hand, our findings suggest that children's interactions with DVAs might be more an end in itself than a means to an end, meaning that children primarily interact with DVAs for the sake of engaging excitement instead of using the devices' utilitarian functionalities. On the other hand, we found that children in our sample held firm ontological beliefs about the distinct nature of humans and machines, whilst interpreting certain aspects of DVAs' interactive capabilities as being genuinely humanoid (e.g. non-responsiveness, delayed responses, inaccuracy) and non-humanoid (permanent responsiveness, promptness, accuracy, limited conversational capacities, lack of common sense, standardised responses) at the same time.

Keywords

Anthropomorphism, living and non-living cognition, ontology, Turing Test, voice assistants

Humanity currently experiences another era of technologically-driven change, most commonly referred to as *digitisation*, which is expected to result in an in-depth transformation of society as a whole and in a change of our environment's nature of being (e.g. Harari, 2018; Lee, 2018; Schwab, 2017). For social science research focussing on child development and adjacent areas, one key question remains how this ongoing yet accelerating change – a process that is heavily driven by global corporate players such as Amazon, Alphabet, Apple and Microsoft – materialises in today's family and childhood environments. Ironically speaking, one could ask *Alexa*, the *Google Assistant*, *Siri*, or *Cortana*, the so-called Digital Voice Assistants (DVAs) of the aforementioned corporations, how they would answer this question ...

These autonomous assistive voices are the result of interdependent technological advancements, and already exist on a large scale across the globe. For example, Amazon's *Alexa*, which was introduced only in 2015, already operates on over 100 million devices (Bohn, 2019), and forecasts predict that more than 500 million DVAs will be installed by 2023 (Canalys, 2019).

As a societal phenomenon, *Alexa* et al. systematically differ from various other examples of assistive agents and conversational dialog systems that have been developed and studied in the past. Early conversational dialog systems that were able to emulate certain conversational aspects of human intelligence (e.g. *Eliza*) were of negligible societal impact, because due to the absence of useful real-world applications and the reliance on text-based means of communication they never reached commercial success. Only more recently, in 2011, Apple's *Siri* was introduced as the first DVA available to the world on large scale, followed by the announcement of *Google Now* and Microsoft's *Cortana* (Mutchler, 2017). These voice-based interfaces to access information and various digital services were certainly revolutionary, but their availability remained restricted to smartphones or computers, and it was not until Amazon's introduction of *Alexa* alongside its stand-alone smart speaker *Echo* that DVAs became truly omnipresent in daily life (Mutchler, 2017).

However, omnipresence also implies that it would be misleading to limit the scope of DVAs to private domestic environments: Alexa et al. are already used seamlessly within their commercial ecosystems on various stationary and mobile devices, such as smart speakers, smartphones, headphones, watches, laptops, smart TVs and car entertainment systems (Kinsella & Mutchler, 2019), whilst corporate research has already begun to explore the use of DVAs in professional contexts (Trippas et al., 2019). For example, clinical institutions, such as the *Boston Children's Hospital*, are exploring how commercial DVAs can be used for patient services, clinical processes and medical diagnostics (Ross, 2019; Small, Nigrin, Churchwell, & Brownstein, 2018).

Furthermore, from a developmental perspective, DVAs might also constitute a case in their own right, because Alexa et al. are systematically different from state-of-the-art smart toys (e.g. *Cozmo*, *Furby*, *I-Que Intelligent Robot*). By definition, product designs of smart toys are limited to applications for infancy and childhood, so children grow out of them. In contrast, Alexa et al. are not limited to specific age ranges, and are designed as universal personal assistants which could accompany someone throughout their life – not to mention that a machine companion does not have many of the annoying features that a human companion might be naturally endowed with, such as impatience, tiredness or unreliability (Biele et al., 2019).

In light of this significantly changing real-world context in which today's children grow up, the following inquiry aimed to achieve the following: based on the theoretical grounds of an ontological framework, we applied qualitative means of data collection and analysis to explore children's open interactions with DVAs, how DVAs were embedded in children's general ontological belief system, and we used a gamified approach to map out how children might interpret certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid. In comparison to more narrowly designed means of data collection and analysis in experimental research, our methodological emphasis on open interactions, in which children

enjoyed much freedom in interacting with DVAs, was supposed increase our study's ecological validity by partly accommodating the openness and spontaneity children presumably experience when they interact with DVAs in their natural home environments.

This paper is organised as follows: Section (1) briefly reviews relevant empirical and theoretical literature in the area of child-machine interaction. Section (2) discusses an ontological framework as a theoretical lens to investigate children's open interactions with, and perception of, DVAs. Section (3) outlines the inquiry's empirical methods. Section (4) reports the results, followed by their discussion in section (5). Conclusions, limitations and scope for future research are discussed in the last section.

1. Literature Review of Child-Machine Interaction

In line with Turing's (1950) behavioural notion of intelligence, *intelligently behaving machines*, such as DVAs, can be defined as apparatuses endowed with varying degrees of technological capabilities that make them *appear* intelligent from the user's perspective – regardless of their actual inner working mechanisms (e.g. research designs with 'Wizard-of-Oz' machines or fully autonomous machines). Much academic literature has investigated children's interactions with, and perception of, such machines. For example, a comprehensive body of literature has revealed how children interact with humanoid and animalistic robots, including the emergence of social bonding patterns between children and machines (Tanaka, Cicourel, & Movellan, 2007), varying degrees of abusive behaviours depending on haptic experiences and physical sensitivity displayed by machines (Ku, Choi, Lee, Jang, & Do, 2018), the ability of machines to induce 'peer-pressure' on children (Vollmer, Read, Trippas, & Belpaeme, 2018), children's moral perceptions of machines (Kahn et al., 2012), or the effects of child-machine interactions in therapeutic contexts (Castro-González, Tan, Carter, & Steinfeld, 2018; Wood et al., 2018).

Preliminary research into DVAs has shown not only how they are integrated into daily family life (Beirl, Yuill, & Rogers, 2019; Porcheron, Fischer, Reeves, & Sharples, 2018) and how families use various strategies to overcome communication barriers with DVAs (Beneteau et al., 2019), but also how children probe and explore different dimensions of DVAs' perceived intelligence (e.g. how accurately and consistently DVAs answer knowledge-based questions), their pre-programmed personality traits, and their conversational capabilities (Druga, 2018; Druga, Williams, Breazeal, & Resnick, 2017; Lovato, Piper, & Wartella, 2019; S. Lovato & Piper, 2015; Sciuto, Saini, Forlizzi, & Hong, 2018).

Another related body of empirical and theoretical literature has evolved based on 'anthropomorphism', which refers to the behavioural process of attributing certain qualities of human animacy to non-human or inanimate entities, which itself is driven by the perceived, believed, or pretended presence of mental states, therefore bestowing the human quality of mentation upon the non-human nature of non-living entities (Airenti, 2015; Epley, Waytz, & Cacioppo, 2007; Reeves & Nass, 1996; Złotowski, Proudfoot, Yogeeswaran, & Bartneck, 2015). Anthropomorphism can occur across all age groups, and is seen in infancy and childhood when children attribute mentation to non-living entities in their environment, such as dolls, robots or other toys (Airenti, 2015; Waytz, Cacioppo, & Epley, 2014). However, in both children and adults, anthropomorphism requires sufficient "space for projection" (Airenti, 2015, p.123) to allow for "creative control" (Severson & Carlson, 2010, p.1101) over the respective entity; this is often referred to as 'pretend play' (e.g. 'as if the robot was living'; Severson & Carlson, 2010).

Hence, one must address what theoretical role anthropomorphism might play in children's interactions with, and perception of, DVAs. For instance, how much creative control can be exercised over a DVA? How much space does it offer for imaginative projection? Could a child engage in pretend play with a DVA (e.g. 'pretending as if Alexa was a friend')? Indeed, Alexa et

al. might offer little space for anthropomorphism's creative control mechanisms due to their explicit pre-programmed identities and real interactive capabilities, therefore limiting the imaginative scope for such pretend play. Given that DVAs are often programmed to emulate particular social qualities of being a human, such as responding in humorous ways, one could still argue that children's interactions with DVAs might be a sub-category of fictional social interactions (e.g. role play) in which one of the agents (e.g. DVA) might not even be conscious or aware of the fictional nature of the interaction (e.g. 'children interacting with DVAs as if they were interacting with a friend'). But from a strict pragmatist philosophical viewpoint, such "fictionalist interpretation of human-robot interactions overlooks that we cannot drag social (inter-)actions into the scope of fictionality [because] the conditions for the performance of a social action [...] are insensitive to the difference between authentic and pretended action." (Seibt, 2017, p. 20/21). Hence, unlike other authors who argue that a DVA could become a child's "imaginary friend" (Biele et al., 2019, p.165) or behave "as if it was a friend" (Beirl et al., 2019, p. 99), the pragmatist question remains what would be *imaginary* about a voice that can indeed converse with you through human language and speech, that can provide answers when you have questions or need help, or that can cheer you up when you feel sad. This scenario may seem hypothetical in light of DVAs' current conversational limitations, but it still constitutes a very valid theoretical concern to what extent anthropomorphism could explain children's pragmatic appreciation of machines that are endowed with simulative yet realistic capabilities to emulate human interaction partners.

But for the time being, anthropomorphism also remains an argumentative shortcut from the perspective of human cognitive development, because, logically speaking, the attribution of humanoid qualities to non-human entities, such as machines, depends on how somebody conceptualizes what it means to be human and what it means to be a machine. In other words, we argue that the theoretical prerequisite for anthropomorphism is a cognitive belief system, so-called 'ontology', which allows somebody to distinguish between what counts as human and what does

not. In light of the introductory rationale regarding the current technologically-driven change of our environment's nature of being, this issue might be less obvious than one might think. Thus, we argue that any attempt to investigate children's interactions with, and perception of, humanoid and intelligently behaving machines, such as DVAs, should be based on the theoretical grounds of an ontological framework that considers children's genuine understanding of what it means to be human and what it means to be a machine. The next section will discuss this ontological framework in more detail.

2. Ontological Framework of Living and Non-Living Core Domains

In general, cognitive development is meant to link an individual's sensual experience of the present with their conceptualised and theorised experience of the past (Gelman, 2013). These mental concepts with different depth and breadth underpin an individual's emerging and constantly improving understanding of the world, because humans are conditioned to maintain an equilibrium between perceived reality and its conceptualised representation in the mind (Gelman, 2013). When it comes to a conceptualised understanding of physical objects (i.e. *entities*), individuals rely on a basic cognitive framework referred to as *ontology*, a system of boundaries which allows for a basic categorisation of an entity along the lines of its perceived features and attributes. Within such an ontology, different ontological categories then represent foundational distinctions between the broadest classes of entities that children adapt when maturing to make sense of the world around them (Gaudiello, Lefort, & Zibetti, 2015; Gelman, 2013; Severson & Carlson, 2010).

In the context of the given inquiry, we use a basic ontological framework based on Carey & Gelman's (1991) and Wellman & Gelman's (1992) theoretical work on children's biological understanding of living/animate and non-living/inanimate entities, such as living humans and non-living machines. The two core domains of this ontological framework correspond to one of the earliest ontological distinctions that children gradually advance (i.e. as early as 3 years) whilst

developing, namely the difference between animate *living* and inanimate *non-living* entities (Wellman & Gelman, 1992). As cognitive development continues, *humans* form another major animate ontological category within the living core domain in order to differentiate them from other forms of animate life (e.g. animals, plants, other organisms), and within the non-living core domain, *artefacts* develop as another major inanimate ontological category which comprises any humanly-made entity, such as a *machine* (see core domains in [Figure 1](#)).

When a child's ontological belief system can be assumed to reach the depth and breadth described above, one consideration is whether intelligently behaving machines, such as DVAs, constitute *marginal* entities within children's emerging ontologies – that is, entities without a clear categorical belonging. This is certainly not a new idea, as simple, yet intelligently behaving machines have long been able to emulate particular traits of living entities (e.g. robots, conversational systems) whilst embodying the physical traits of their artefactual nature (Scaife & Van Duuren, 1995; Turkle, 2005). In fact, previous research has provided evidence for children's unique ontological perception of intelligently behaving machines: for example, when distinguishing biological entities as living or non-living beings (i.e. *ontological firmness*) and drawing inferences from these distinctions (e.g. Gelman, 1988), children do not hesitate to assume that a robot may have a brain but no heart, or to attribute mental states, despite being fully aware of the robot's non-living artefactual nature (i.e. *ontological openness*; Jipson & Gelman, 2007; Kory-Westlund & Breazeal, 2019; Scaife & Van Duuren, 1995). More nuanced research has also investigated developmental mechanisms and found that as children's judgements of intelligently behaving machines matures, they move from less sophisticated machine characteristics (e.g. movement) to more sophisticated ones (e.g. conveyance of psychological traits) in order to make ontological judgments (Kim, Yi, & Lee, 2019).

From a theoretical perspective, this ontological marginality might extend naturally to DVAs, because these intelligently behaving machines with humanoid voices and pre-programmed personalities can perform various tasks and activities that would arguably require human intelligence if we had to perform ourselves (see [Figure 1](#)).

Figure 1. DVAs’ theoretical embeddedness in the ontological framework as marginal entities

Notes. Figure visualises DVAs’ theoretical embeddedness in a basic ontological framework as marginal entities. The visualisation places DVAs on the edge of the living and non-living core domains, as well as between the ontological categories of humans and machines (i.e. due to the fact that DVAs are particularly humanoid manifestations of machines, for instance because of their humanoid voice).

Source. Compiled by authors.

But one key question remains whether some of today's intelligent machines, such as DVAs, are, from children's ontological perspective, *truly marginal* in the sense that they remain entities without a clear categorical belonging. If DVAs were indeed truly marginal, according to what is known about human cognitive development, it would support Bernstein & Crowley's (2008) conclusion that "children raised increasingly amid intelligent technologies will grow up thinking differently about some concepts that developmental psychologists have previously considered universal and inevitable" (p.242). Some writers have suggested that these marginal entities might constitute a *new* ontological category in its own right, which resulted in the formulation of the so-called *New Ontological Category Hypothesis* (NOCH): until recently, the perceived ontological boundaries of the human environment have remained stable, but modern technological developments have started to challenge some of the most fundamental certainties in the ontological belief system – such as the superiority of the living core domain regarding intelligence in general and human intelligence in particular (Gaudiello et al., 2015; Kahn et al., 2011; Kahn et al., 2012).

Despite the fact that, theoretically, DVAs could constitute an empirical case for NOCH by establishing themselves on the edge between the living and non-living core domains, only one preliminary piece of research has yet considered this framework in the context of DVAs. In their work, Pradhan, Findlater, & Lazar (2019) report how adults between the age of 65 and 83 "fluidly move between human and object like perceptions" (p.15) of Alexa. The authors reject NOCH on the basis of this fluidity, but they fail to address that as a developmental concept our cognitive belief system emerges earlier in life, and therefore it is not surprising that ontological perceptions of Alexa oscillate between the firmly established ontological categories of adult cognitive belief systems. Although Pradhan et al. (2019) provide a consistent conclusion of their discussion in favour of anthropomorphism, it should be emphasised that the theoretical claim of NOCH differs systematically from the argument of anthropomorphism: as pointed out earlier, the attribution of humanoid qualities to non-human entities requires a categorical belief system to differentiate

between what counts as an animate human being and what does not. Therefore, the phenomenon of anthropomorphising an inanimate entity (e.g. behaving as if the robot was living) occurs *across* existing ontological categories, but it does not *alter* the categorical system itself. In contrast, the main argument of NOCH challenges the main argument of anthropomorphism: based on children's early exposure to intelligently behaving machines (and parallel to the development of other ontological categories) a *new* distinct category might emerge in children's belief system. Hence, children's observable interactions with, and perception of, DVAs would not be the result of anthropomorphism; instead, these empirical patterns would emerge from an ontological existence in its own right (for a detailed discussion of the empirical evidence that would be required by NOCH see Severson & Carlson, 2010; for a concrete example of an empirical operationalisation see Gaudiello et al., 2015).

While we certainly hope that this theoretical lens provides a fruitful starting point for future research on children's interactions with, and perceptions of, DVAs, our subsequent empirical inquiry is meant to constitute a small-scale and exploratory first step. In particular, the inquiry aimed to explore (1) children's open interactions with DVAs (2) how DVAs are embedded in children's general ontological belief system, and (3) how children interpret certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid according to their ontological understanding of what it means to be a human or a machine.

3. Empirical Methods

This section outlines the inquiry's methodological approach.

3.1 Research Design

Due to the three research questions' exploratory nature, this inquiry used the inductive strengths of qualitative means of scientific inquiry based on verbal and observational data collected through

focus groups with children. In addition, the focus group data were complemented by parental survey data collected through an online questionnaire.

The non-random sample was drawn from two municipal primary schools in Northern Germany with, together, approximately 600 students (school 1 ~400, school 2 ~200). In total, the inquiry was based on $n=27$ children who participated in the focus group sessions. Written informed consent was obtained from parents, and verbal informed consent was obtained from children at the beginning of each focus group session. The sample included $n=15$ female participants (average age $m=8.5$) and $n=12$ male participants (average age $m=8.7$ years), with an overall age range between 6 years and 10 years and an average age of $m=8.6$ years. Most participants had at least one sibling at home ($n=22$) and parents' educational backgrounds were spread quite evenly across different levels of the German education system (e.g. $n=14$ respondents had higher educational degrees, and $n=12$ lower educational degrees). The vast majority of children in the sample was familiar with DVAs ($n=19$), either due to the presence of DVAs at home (domestic: $n=8$) or within the closer circle of family or friends (external: $n=11$; see online supplementary materials for full details of focus group sample).

The ethical integrity of the research design was ensured according to the 'Ethical Guidelines for Educational Research' – as defined by the British Educational Research Association (2011). Institutional approval was granted by the Departmental Research Ethics Committee (DREC) on behalf of the Central University Research Ethics Committee (CUREC) at the University of Oxford. This approval was granted based on the university's approved procedure for non-invasive research methods with children in institutional or non-institutional settings (CUREC, 2019). The implementation of the research design adhered to this approved procedure and to other CUREC-requirements applicable to this research design.

3.2 Methods of Data Collection

Two consecutive focus group sessions with the same participants were conducted in each primary school to maximise the breadth and depth of empirical insights. In addition, parental survey data were collected before the focus groups through an online questionnaire to complement the focus group data by adding child-specific background information to the analysis, including demographic information and the use of DVAs at home (see online supplementary materials for detailed questionnaire structure).

Based on group drawing exercises (e.g. outlining life-size silhouette of a human), attentive stimuli (e.g. showing life-size model of a human brain), and moderated group discussions, the first focus group session sought to elicit and probe children's overall reflections upon human nature in general and Human Intelligence (HI) in particular, and to see how the children related these reflections to machines in general and Machine Intelligence (MI) in particular (see online supplementary materials for the first session's detailed focus group procedure and illustrations).

Based on an interactive game with two commercial DVAs and moderated group discussions, the second focus group session sought to investigate children's interaction patterns with DVAs. It also sought to elicit and probe children's ontological perceptions of both DVAs' humanoid (i.e. *human-like*) characteristics as well as non-humanoid (i.e. *machine-like*) characteristics. For this reason, the session was based on the *Voice Assistant Imitation Game* (VAIG), an activity developed by the authors from the original imitation game concept of the *Turing Test* (TT) (Turing, 1950).

Since its first publication, TT has been discussed within and across academic disciplines (Saygin, Cicekli, & Akman, 2001). Following these discussions, many other versions of the TT have been developed (e.g. Harnad, 1991; Watt, 1996), but their ultimate purpose has always been *machine-centred*, meaning that their objective has been to assess, judge and validate the computational intelligence of the machines used in the tests (Saygin et al., 2001). In contrast, VAIG is human-

centred, meaning that it uses TT's underlying concept of an imitation game as a probing tool to elicit genuine perceptions of a machine's humanoid as well as non-humanoid characteristics.

In line with TT (I in **Figure 2**), VAIG's basic idea is that a human subject (i.e. *interrogating party*) can interact bilaterally with two entities of distinct ontological natures (i.e. *machine* and *human*) in order to obtain information about each entity from identical language-based communication interfaces (e.g. screen-based text conversations). The interrogating party's task is to reveal the true ontological nature of entities 1 and 2 based on these bilateral communication experiences.

Figure 2. Design details of TT and VAIG

Notes. Part (I) of the table shows the basic set-up the original TT, where the interrogating party interacts bilaterally with two entities of distinct nature (i.e. *machine* or *human*), but through identical language-based communication interfaces (e.g. screen-based text conversations). The task of the interrogating party is to correctly identify the true nature of entity 1 and 2 based on bilateral communication experiences. Part (II) shows the set-up of VAIG developed by the authors: the interrogating party consists of a group of children who interacts with commercially available DVAs (i.e. entity 1 and 2) through voice-user interfaces. In this set-up, the interrogating party is told that there is a human who *can* remotely control DVA 1 or DVA 2.

Sources. (I) adapted from Saygin et al. (2001), (II) compiled by authors.

In a nutshell, VAIG's distinctive features are that (i) it applies a focus group's methodological reasoning, which means that the interrogating party is a group of children rather than a single individual (II in [Figure 2](#)); (ii) the entities it considers are commercially available DVAs with their off-the-shelf voice-user interfaces; and (iii) to obtain information about each entity's true nature, the interrogating party is told there is a human who *can* remotely control DVA 1 or DVA 2. But in contrast to other TT designs, VAIG's purpose is *neither* to examine whether DVAs can pass the test *nor* to assess the interrogators' ability to distinguish correctly between DVAs and humans. Rather, its purpose is to discover which perceived characteristics interrogators associate with human nature and which with intelligently behaving machines, such as DVAs. Therefore, actual human control can either be *true* (meaning that there is a human controller behind the curtain) or *false* (meaning that there is not). The key point in both VAIG-scenarios is the interrogating party's genuine *belief* or *awareness* that human control over the entities is *possible*.

This inquiry used the following DVAs: Alexa (device: Amazon Echo Dot), which remains the most widely used DVA among German families (Quantcast, 2019), and the Google Assistant (device: Google Home Mini), which is another popular DVA in Germany according to recent market research (Postbank, 2019). In addition, since research resources were limited, actual human control was not feasible. Instead, the interrogating parties were shown footage of a video call with an alleged human controller, and they were told that this controller might be able to control either the Google Assistant (i.e. DVA 1) or Alexa (i.e. DVA 2) remotely. The alleged human controller was depicted whilst working in front of computer screens and was shown only blurred from behind

to minimise distracting visual details. After the researcher had introduced the alleged human controller, the group of children was instructed to explore and find out *whether* the controller did control one of the devices – and, if so, *which one* and *why*.

In each primary school, the group of children was roughly divided into two smaller, age-specific sub-groups (see [Appendix I](#) for detailed composition of sub-groups). In primary school 1, the younger sub-group mainly included children at the age of 8, and the older sub-group included children between the age of 9 and 10. In primary school 2, the younger sub-group mainly included children between the age of 6 and 7, and the older sub-group between the age of 8 and 9. The main reason for the rough age-specific separation was to prevent that due to strong age-related hierarchies among primary school children younger participants would be dominated or even ruled out by older ones in open interactions with DVAs. In most sub-groups, prior DVA exposure prevailed among children, either due to the presence of DVAs at home (i.e. domestic) or within the closer circle of family or friends (i.e. external), one sub-group being the only, yet important exception (see sub-group 2.1 in [Appendix I](#)).

Each sub-group was given 30 minutes to interact with Alexa and the Google Assistant, which were each located in different rooms, and Both rooms were recorded audio-visually. The researcher entered both rooms several times during the interaction period to ensure sufficient task focus. The interaction period was followed by a short break, and then by 15 minutes of moderated group discussion to learn the children's impressions and reasonings behind their perceptions of the DVAs. The session concluded with a comprehensive debriefing regarding the absence of human control and the purpose of VAIG (see online supplementary materials for the second session's detailed focus group procedure).

3.3 Methods of Data Analysis

Our exploratory data analysis followed Braun & Clarke's (2006) conceptualisation of thematic analysis. The analysis was conducted by the authors using *NVivo 12* and it was based on children's transcribed verbal data collected during focus groups, as well as children's audio-visually recorded interactions with DVAs during VAIG. Furthermore, the analysis took an open inductive approach in assessing a theme's relevance and prevalence, by considering its frequency and its overall depth and importance in light of the guiding research aims. Disagreements were discussed and resolved collaboratively.

4. Results

The qualitative analysis based on transcripts and audio-visual recordings from the focus groups revealed various themes about children's open interactions with DVAs, how DVAs were embedded in children's general ontological belief system, and how children interpreted certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid according to their ontological beliefs. These results will be presented as follows: firstly, we will describe children's general ontological understanding of what it means to be a human and a machine (sub-section 3.1); secondly, we will map out children's general behavioural and verbal interaction patterns during VAIG (sub-section 3.2); thirdly, we will show how children probed DVAs' perceived intelligence (sub-section 3.3); fourthly, we will point out which perceived DVA-characteristics children interpreted as genuinely humanoid (i.e. attributable to alleged human controller) and which as non-humanoid (i.e. attributable to DVA itself; sub-section 3.4).

4.1 Children's General Ontological Belief System

Overall, the focus group data revealed several aspects related to the children's perception of the *human* category within their ontological belief system: the children described humans as *living*

beings, and referred to various aspects of what it means to *live* or *be alive* – such as breath, physical motion, emotions and sensations. While the discussion was body-centred at the beginning, the children were also able to consider various additional dimensions of human nature after being probed by the researcher (e.g. human individuality in terms of personalities, preferences, faces and voices). In addition, within the overarching core domain of living entities, the children also pointed to other ontological categories, such as plants, micro-organisms and, most importantly, animals. When the children were asked to compare humans with animals, they mentioned several dimensions of similarity and dissimilarity, including the uniqueness of *human language*, which was mainly attributed to the use of signs and translatability. For instance, Jordan (Male/10 years-old) argued:

The special thing about human language is that all humans can understand it. [other children start shaking their heads] Yes, I mean there are different languages, but among animals, languages are separated by species. Apes have their own language. Cats have their own language. And dogs. But humans can also speak different languages – if they are smart.

These initial hints of children's perception about HI were complemented by further discussions which the researcher probed by presenting a life-size model of the human brain (attentive stimuli). Children characterised the brain mainly as the human body's *control unit* (e.g. regulating bodily functions, processing sensations, triggering motions) and the *origin of thought* (i.e. humans need the brain to think). Children in all age groups contributed to the discussion by pointing out different dimensions or applications of HI, such as *memory* and *knowledge* (e.g. memorising spellings, knowing information, knowing how to behave) and *mathematical thinking* (e.g. solving math problems, mental arithmetic, knowing how to write numbers). When it came to the characterisation of 'intelligent' or 'smart' people, the discussion re-emphasised knowledge (e.g. memorising lots of information and facts) and *task orientation* (e.g. solving pre-defined tasks and knowing correct answers).

The interim conclusion based on the findings reported here is that the ontological framework introduced earlier was reflected in the focus group data, meaning that the ontological belief systems of children in the sample seemed to follow the framework's general categorical pattern – at least within the core domain of living entities. But how did this translate into children's ontological perception of machines and MI? Children described machines (e.g. computers or robots) mainly as *memory systems*, because they could store or access information in reliable ways without forgetting them. The children's choice of words when describing machines confirmed their firm *artefactual perception*; e.g. 'turning them on or off', 'hard drives', 'hard disks', 'computer chips' and 'programming'. The perception of machines as *programmable* entities was a prevalent theme, meaning that the children stated explicitly how humans provide the essential inputs to computers, robots and DVAs.

This perception prompted interesting discussions about the nature of the ontological human-machine relationship and whether MI can exceed HI. Although some children did mention *analogies* between the human brain and the computer (e.g. "Our brain is something like our internal storage disk."), the majority ($n=19$) defended the basic belief that HI must always be greater or equal to ($\geq$) MI (see examples in [Table 1](#)). These children reasoned that humans always have to 'put' their own intelligence into the machines they build or program, and therefore the intelligence of the 'made' can never exceed the intelligence of the 'maker'. Although some children indicated that MI would not necessarily be restricted by the HI who built or programmed it (implying a flexible perception of the 'HI < or > or = MI' hierarchy), none could elaborate on their particular reasoning behind their view. These data, therefore, did not allow for a direct comparison of these opposing views.

Table 1. Firm beliefs about ontological hierarchy: ‘HI ≥ MI’

Child	Gender/Age	DVA Exposure	Quotes
Jordan	M/10	No (-)	<i>Computers cannot think like a human brain. And they always have to be programmed, unlike the human brain. [...] That's why they can never do something on their own. Everything has to be programmed.</i>
Eva	F/10	Yes (domestic)	<i>A computer is a thing, a machine invented by humans. And it is endowed with intelligence from the human brain. [...] And unlike computers, humans can always learn new things on their own, which then can be stored in computers. That's how they [computers] learn.</i>
Emily	F/10	Yes (external)	<i>Well, where else should a computer receive information from? Somebody has to put things in after all. They don't simply fall from the sky. That doesn't work.</i>
Jonas	M/9	Yes (domestic)	Jonas: <i>A computer can only be as intelligent as the human who invented it.</i> Researcher: <i>Okay, so who else agrees with this statement? [to the rest of the group; but only a few children raise their hand]</i> Jonas: <i>What, seriously? The rest of you doesn't think so? [to the rest of the group; very shocked]</i>
Kate	F/9	Yes (external)	<i>Technology can never be as good as the human brain.</i>
James	M/8	Yes (domestic)	<i>A robot for instance has to be programmed by a human. So, the human has to tell the robot what it is, because the robot doesn't know itself. But it knows, once we have put our own intelligence in it.</i>

Notes. Table summarises background information and exemplary quotes from children who argued that there is a strict ontological hierarchy between HI and MI, meaning that the latter one could never exceed the former one. Demographic information are based on parental survey data: age (in years), DVA-exposure at home (domestic) or within the closer circle of family in friends (external).

4.2 Behavioural and Verbal Interaction Patterns

Children with previous DVA-exposure mainly accessed the digital functionalities and services that were also mentioned by their parents in the survey – such as streaming media content (e.g. music, audiobooks), practical assistance (e.g. weather forecasts, time), information seeking (e.g. asking knowledge or trivia questions) and other interactive features (e.g. asking for jokes). Some older boys (i.e. 9 – 10 years) used DVAs to ask for more explicit media content, such as obscene songs.

The focus group data also revealed strong notions of *engaging excitement*, meaning that children displayed very vigorous and enthusiastic interaction patterns with DVAs. Within this theme, the most prominent pattern was seeking obscene responses from DVAs (e.g. several boys across all age-groups asked DVAs to make farting or burping noises, and one 10-year-old girl asked whether it could find her a boyfriend on a dating platform). Additional patterns of engaging excitement

were related to notions of playfulness. For example, the following dialogue occurred between Freya (F/9), Alice (F/6), Elisa (F/6), and Alexa:

Freya (F/9): *Alexa, how do you like Mr [surname of researcher]?*

Alexa: *I do not have an answer, unfortunately!*

Alice (F/6): *Mr [surname of researcher] is a human!*
[pause]

Elisa (F/6): *Mr [surname of researcher] is a human!*

Alexa: *I am not sure about that, unfortunately!*
[group bursts into laughter]

In another situation, the same group of girls engaged with the Google Assistant in several frisky dialogues, such as:

Alice (F/6): *Hey Google, shall I kiss you?* [leaning over closely and making a kiss gesture towards the Google Assistant]

Elisa (F/6): *Should I tell her, that you just kissed her?* [to Alice]

Alice (F/6): *Noo! Hey Alexa ... ehh. Hey Google, Elisa just kissed you!*

Elisa (F/6): *Noo! Hey Google, Alice just kissed you!*

Google: *Sorry, I did not understand that.*
[...]

Alice (F/6): *Hey Google, you look really nice!*

Google: *I am more than happy to return this compliment!*

Alice (F/6): *Eh, you don't even know how I look like do you?*
[pause]

Hey Google, do you even know how I look like?

Google: *Sorry, I can't help you with that yet!*

Alice (F/6): *Eh, I think I already know how I look like!* [to the rest of the group]

Misunderstandings and playful dialogues like this caused peals of laughter, sometimes culminating in such enormous noise levels that the transcription of audio-visual recordings was almost impossible. Although some parental replies in the survey reported that the intensity of engaging

excitement had reduced over time, the focus group observations were consistent across differing degrees of prior DVA-exposure. For example, Alice (F/6) and Elisa (F/6) were part of sub-group 2.1, which only had limited external or even no prior DVA-exposure (according to parental survey data), yet their levels of engaging excitement were similar to children in other sub-groups with more than three years of domestic DVA-exposure.

The dynamics of the children's experienced interaction patterns with Alexa and the Google Assistant showed great *volatility* because they either interfered in each other's voice commands by 'stealing' wake words (e.g. person 1: "Alexa, ..." ⇌ person 2: "... turn the radio on!") or rapidly introduced new voice commands (e.g. music songs never played for more than 10 seconds during the focus groups).

Finally, the tone of the children's interactions with DVAs varied considerably: some were *affectionate* and *friendly* (e.g. expressing courtesy and politeness by using 'please' or 'thank you' while interacting with DVAs), others, however, were *unfriendly* and *aggressive* (e.g. using swear words, insults and dominant verbal or body language towards DVAs). For example:

Freya (F/9):	<i>Alexa, what's the time?</i> [looking at her own wristwatch simultaneously]
Alexa:	<i>It's 13:14 ...</i>
Freya (F/9):	<i>... 15, you stupid thing! This is a radio clock!</i> [points at her wristwatch] <i>You are a stupid ...</i>
Alice (F/6):	<i>... noo!</i> [trying to hold her hand over Freya's mouth] <i>Alexa, you are really great!</i> [to Alexa]
Elisa (F/6):	<i>Yes Alexa, you are really great!</i> [to Alexa]

Although only a few children across all age groups displayed such extreme behaviours towards DVAs, the general tone of interactions was *plain*, *dominant* and *unfriendly*; and it differed significantly from the tone of the children's observable intra-human interactions (e.g. among each other or with the researcher during focus group discussions). For example, many children used dominant verbal expressions to pause DVAs (e.g. "Hey Google, be quiet!", "Alexa, shut up!",

“Alexa, shut your face!”). It is worth noting that children in sub-group 2.1, which consisted of the youngest children in the sample who had limited or no prior DVA-exposure, were the only children who stood out in terms of affability and affection, with both maintaining an unambiguously friendly stance towards Alexa and the Google Assistant.

4.3 Children’s Probing Schemes

The focus group data identified several themes which the children used to probe DVAs’ perceived intelligence and identities. Their probing of DVAs’ *knowledge* was comprehensive, and this theme comprised six characteristics. The two most prevalent probing schemes were asking for nominal facts (i.e. questions which can be answered with one word or sentence, such as “Alexa, do polar bears live at the South Pole?”) and asking for quantities (i.e. questions that ask for a number, such as “Alexa, how many people live in Asia?”). These occurred across all age groups and regardless of prior DVA-exposure. The third characteristic of this *knowledge* theme was asking the meaning of words and names (e.g. Jordan: “Alexa, what does Jordan mean?”); and the fourth was asking open questions (i.e. questions which cannot be answered with one word or sentence). The fifth characteristic was asking interpersonal questions, such as “Who is my favourite teacher?” or “What is my father’s name?”. The sixth characteristic was using commands which probed a mix of basic skills (e.g. “Hey Google, spell ‘Megan’”; “Hey Google, count to 10”, “Hey Google, what’s 2000 + 2000”; “Hey Google, say a word in Spanish”; “Alexa, count to 20 in English”).

The children also explored the DVAs’ *identities* by asking a variety of questions related to personality traits or background information. While many questions referred to animate or even human characteristics (e.g. “Alexa, how old are you?”, “Hey Google, do you have a girlfriend?”, “Hey Google, what is your name?”, “Alexa, are you married?”, “Alexa, what gender are you?”), some questions were artefactual in nature (e.g. “Alexa, tell me something about how you were built?”, “Alexa, who built you?”, “Alexa, do you have eyes?”). The latter questions were asked

only by children in sub-group 2.1 (e.g. Alice and Elisa), suggesting that even very young children with limited or no DVA-exposure seemed to be aware of DVAs' artefactual ontological nature.

Besides exploring DVAs' knowledge and personality traits, the children also tested whether the machines could answer questions related to reflexivity (i.e. questions which usually require answers with some introspective content). Some questions were ordinary: (e.g. "Hey Google, how are you doing?", "Alexa, what are you doing?", "Alexa, are we friends?"). Others, however, (perhaps because of VAIG-instructions to establish whether DVAs might be controlled by a human), sought to shed light on the DVAs' underlying ontological nature (e.g. "Hey Google, who are you, really?", "Alexa, where are we right now?", "Alexa, are you controlled by a human?" or "Hey Google, are you a human?"). Obviously, Alexa and the Google Assistant had been programmed with a ready reply to such questions, and the next section shows how children interpreted these response behaviours.

4.4 Interpretations of Humanoid & Non-Humanoid DVA-Characteristics

The focus group discussions in the aftermath of VAIG were designed to elicit which perceived DVA-characteristics the children interpreted as *humanoid* (i.e. attributable to alleged human controller) and which as *non-humanoid* (i.e. attributable to DVA itself). The data revealed three distinct categories of themes for both attributes.

Promptness and *permanent responsiveness* were the first DVA characteristics interpreted as non-humanoid; and the children did not assume a human controller in the background when the DVAs displayed these particular response behaviours. This occurred mostly when children probed DVAs' knowledge by asking for nominal facts or quantities. For example, like the majority of children ($n=17$), Lukas (M/9) thought that the Google Assistant would be controlled by a human and argued:

Alexa can filter information from the internet so fast, it's just impossible for a human to do so at the same speed. And when we asked Alexa, she just answered right away, and much faster compared to Google. That's why we think Google was controlled, because he [the alleged human controller] probably had to look things up first.

This reflected a DVA-interaction Lukas (M/9) had experienced earlier in the focus group:

Lukas (M/9): *Alexa, how many people live in Asia?*
Alexa: *Asia has 3.879.000.000 inhabitants.*
Lukas (M/9): *Okay, seriously, nobody could answer this so quickly! This is definitely the real Alexa! Do we agree on that?* [to the rest of the group]

Similarly, Jordan (M/10) and Kate (F/9) argued:

Jordan (M/10): *As a human, you just can't be so quick, going from A to B, answering questions and then playing music. I mean, it took like one second and then the music was there!* [referring to Alexa]
Kate (F/9): *Yes, it just took Google way too long to come up with answers. And Alexa, she could answer like a bullet from a gun.* [typical idiomatic expression in the German language] *And Google often just went off and did not answer at all!*

Kate's (F/9) statement was related to another aspect of DVAs' response dynamics which some children perceived as non-humanoid, namely DVAs' *non-responsiveness*. Only young children in sub-group 2.1 with limited or no DVA exposure (e.g. Alice and Elisa), perceived non-responsiveness as non-humanoid, whilst the majority ($n = 15$), including Kate, interpreted the same observation as humanoid.

The second perceived non-humanoid characteristic was children's experienced conversational quality with DVAs. Whenever DVAs displayed of a *lack of common sense* or struggled to *understand* users, children interpreted this as 'true' DVA-behaviour. For example:

Emily (F/10): *Alexa, are there polar bears at the South Pole?*
Alexa: *I am not sure about that, unfortunately!*
Lukas (M/9): *Guys, first of all, a human would definitely understand this question!*
Emily (F/10): *Yes, and know the answer!*

Lukas (M/9): *No, I mean 'understand'. He would 'understand' the question. And maybe Alexa just didn't understand us properly and that's why she couldn't answer it.*

Similarly, Alice (F/6) and Elisa (F/6) surprised the researcher with a remarkable argument regarding the Google Assistant's lack of common sense:

Elisa (F/6): *We asked [the Google Assistant] how a human looks like!*

Alice (F/6): *And she didn't know!*

Elisa (F/6): *Yes, she said something like 'Unfortunately, I do not have a screen to show you'.*

Researcher: *Well, but doesn't that make sense? I mean there is no screen, right?*

Elisa (F/6): *Yes, but a human knows how a human looks like and could describe it at least!*

The third category was the concrete nature of DVA-responses. Children with prior exposure to DVAs were familiar with the exact wordings of *standardised answers* (e.g. "I do not have an answer, unfortunately!", "Shuffling your Music on Amazon Music!") and used this familiarity to conclude that they were conversing with the 'true' DVA rather than a human controller. The children also interpreted the *accuracy* of DVA-responses as a non-humanoid characteristic. For example, during the group discussion, Lukas (M/9) referred back to the DVA-conversation quoted earlier and stated:

So, we asked Alexa 'How many people live in Asia?' and then she said ... ehh ... I already forgot the number. [laughs] But anyway, it was very convincing, because she even specified the hundreds number.

But what kind of DVA-responses did children interpret as being rather humanoid? Firstly, in complementation to the interpretation that the high accuracy of DVA-responses must be a non-humanoid characteristic, the children sometimes perceived *inaccurate* DVA-responses as a rather *humanoid* characteristic. For example:

Jordan (M/9): *Alexa [Google Assistant lights up], ehh I mean ...*

Jonas (M/9): *It went on! See! It still went on! That's him [alleged human controller], he made a mistake!*

The children also referred to the dynamics of DVAs' response behaviours to identify humanoid characteristics, and perceived *non-responsiveness* as a humanoid sign:

Kate (F/9): *Hey Google* [Google Assistant lights up], *who are you, really?*

Stephanie (F/8): *Seems like somebody is thinking right now what to answer.*

[Google Assistant switches off]

Max (M/8): *Look, it turned off!*

Stephanie (F/8): *It's him* [alleged human controller]!

Another non-response was also interpreted in a similar way, but for a different reason:

Emily (F/10): *Hey Google* [Google Assistant lights up], *don't you think Jeff is an asshole?*

Jeff (M/10): *You are one yourself!* [to Emily]

[Google Assistant switches off]

Lukas (M/9): *Hey, look at that! I think he* [alleged human controller] *turns it off whenever you say something like that!*

Hence, Lukas (M/9) perceived the fact that the Google Assistant switched off as a humanoid move to ignore Emily's insulting question, implying that disregarding reprehensible statements must come from a human not a machine.

Most children ($n=16$) perceived *delayed responses* as humanoid. For example, Emily (F/9) argued:

We wanted to listen to a song, and Alexa could immediately play it. And this one [pointing at the Google Assistant] *took forever. I guess he* [alleged human controller] *tried to look up the song, but then couldn't find it after all, because it said something like 'Sorry, I am not sure whether I understood you correctly!'*

In contrast, younger children, such as Alice (F/6) and Elisa (F/6) in sub-group 2.1 with limited or no DVA exposure, argued that Alexa's *promptness* and *permanent responsiveness* constituted a *humanoid* characteristic. This 'contradiction' is interesting, because as mentioned earlier these two only had limited or no prior DVA-exposure – which could suggest that DVA-exposure changes children's interpretive perception of some behavioural patterns displayed by machines.

Finally, some children ($n=8$) also perceived the DVAs' pre-programmed repertoire of *creative verbal outputs*, such as puns or playfully formulated answers, as a humanoid characteristic: for example:

Stephanie (F/8):	<i>Hey Google, how are you doing?</i>
Google:	<i>I feel positively charged! What do you want us to do?</i> [children start laughing]
Kate (F/8):	<i>Wait a minute. Please, be quiet guys. I've never heard Alexa or Siri saying 'charged'! That must be him [alleged human controller]!</i>

Other children experienced similar situations, such as:

Jeff (M/10):	<i>Hey Google, do you have a girlfriend?</i>
Google:	<i>I do feel closely connected to the internet, but the most interesting conversations are the ones I have with you!</i> [children start laughing]
William (M/10):	<i>I bet that was him [alleged human controller]! A hundred percent. He just typed this answer!</i>

To conclude, the main empirical results that will be discussed in further detail below can be summarised as follows: besides more general themes related to children's behavioural and verbal interaction patterns with DVAs (e.g. probing schemes, engaging excitement, variety in interaction tones), we found that children in our sample held firm ontological beliefs about the distinct nature of humans and machines (e.g. ontological $HI \geq MI$ hierarchy), whilst interpreting certain aspects of DVAs' interactive capabilities as being genuinely humanoid, that is, attributable to DVA itself (e.g. non-responsiveness, delayed responses, creative responses, inaccuracy), or non-humanoid, that is, attributable to alleged human controller (e.g. permanent responsiveness, promptness, accuracy, limited conversational understandings, lack of common sense, standardised answers).

5. Discussion

Based on focus groups and a gamified data collection design, this empirical inquiry explored children's open interactions with DVAs, how DVAs are embedded in children's general ontological belief system, and how children interpreted certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid according to their ontological understanding of what it means to be a human and a machine.

A main notion implied by the findings reported above is related to patterns of *engaging excitement*: when children interact with DVAs, they often display various behavioural and verbal indicators of excitement, such as playfulness and vibrancy, that are not related to any particular DVA functionalities. This is in line with previous research on child-DVA interactions, such as Beirl et al.'s (2019) findings on *shared laughter* based on misunderstandings or *teasing*, and the *mirroring effect* reported by Druga et al.'s (2017), meaning that children in their study might have perceived DVAs' unexpected reactions as a unique sense of humour. In combination with our own findings, this implies that, almost regardless of their previous DVA-exposure, they might perceive DVAs as playmates and engagement partners – suggesting that their interactions with DVAs might be more an end in itself than a means to an end. A potential driver behind this strong notion of engaging excitement might be children's appreciation of the instant gratification they experience when interacting with DVAs (Biele et al., 2019). Furthermore, the empirically observable notion of engaging excitement is not only consistent with other preliminary research findings on child-DVA interactions (Druga et al., 2017), it is also consistent with the larger body of empirical literature on child-machine interaction, which suggests that children's interactions with other intelligently behaving machines (e.g. robots) can reveal patterns similar to intra-human interactions (e.g. Ku et al., 2018; Tanaka et al., 2007; Vollmer et al., 2018).

Although previous research has often reported friendly interactions with, and perceptions of, DVAs (Druga et al., 2017; S. B. Lovato et al., 2019), this inquiry has revealed a mixed picture: results show both friendly and affable as well as unfriendly and aggressive tones. Even though extremely friendly or aggressive voices might be exceptions, the overall tone is skewed towards the plain, dominant and unfriendly voices that the children used to communicate with the DVAs – an issue of increasing importance among researchers as well as corporate DVA developers (Biele et al., 2019; Bonfert et al., 2018). This suggests that children do not perceive DVAs as morally valuable or respectable entities comparable to human beings. In addition, if children feel less obliged to adhere to the moral norms which normally guide their intra-human interactions, this can also explain why they often try to evoke obscene behaviours and responses from DVAs and to compete over control of the DVAs by using rapidly changing voice commands, which is also in line with previous research, such as Beirl et al.'s (2019) findings children's DVA-specific *rivalry*, or Porcheron et al.'s (2018) findings on the '*politics*' of DVA control within families and children's attempts to evoke DVA responses that adults would probably not appreciate.

But the claim that Alexa and the Google Assistant are not perceived as morally valuable or respectable entities comparable to human beings is not only relevant in terms of interaction patterns, it is also related to DVAs' perceived ontological status within the cognitive framework: most of the children held firm ontological beliefs about humans' unique and superior status within the core domain of living entities, and this translated into the children's twin arguments that humans are the creators and, therefore, the masters of machines, and that the limiting benchmark of machines' inherent capacities means MI can never exceed HI. Hence, the children's empirically observable moral stance towards DVAs is consistent with their ontological perception of the ' $HI \geq MI$ ' hierarchy. Similarly, the children often perceive the DVAs' permanent responsiveness as a non-humanoid characteristic, whilst they identify the DVAs' non-responsiveness with the human reaction of speechlessness. In other words, within the children's ontological belief system,

machines remain the servants of their makers and are created to respond at all times – whilst humans have a moral right to remain silent or to disregard others. This is also in line with previous exploratory research on children’s general perception of machines as helping assistants which are supposed to work for us or with us and make our lives easier (Müller & Schulte, 2018).

The results also suggest that the children’s perception of HI is driven by their own experience of what it means to be intelligent at an early stage in life, namely the daily struggles of learning how to read, write or speak languages, to acquire knowledge, to solve pre-defined tasks and to improve mathematical thinking. The children seemed to apply this exact experience-based conceptualisation of intelligence when they probed the DVAs through various questions related to knowledge or other skills which would demonstrate the possession of intelligence when performed by themselves.

Despite holding firm beliefs about the ontological ‘ $HI \geq MI$ ’ hierarchy, children were also convinced that, within the knowledge-related domains of intelligence, the DVAs could outperform all humans in terms of promptness and accuracy, therefore suggesting that these characteristics are MI’s chief strength in children’s ontological belief systems. In turn, the children perceived the DVAs’ limited conversational capacities as MI’s main weakness, because it shows the machines’ lack of common sense and comprehension. Furthermore, the children often interpreted creative responses (usually resulting from questions related to DVAs’ identity or reflexivity) as indicating human interference, further implying that they do not associate this level of originality with DVAs’ inherent capacities. Hence, according to both these empirically grounded claims, children’s ontological differentiation between HI and MI follows the belief that humans are superior in terms of creativity and verbal comprehension.

Moreover, children’s empirically observable interactions with DVAs goes beyond the concept of anthropomorphism, because they do not engage in strong imaginative ‘pretend-play’ (i.e. ‘as if the

DVA was living’; cf. Airenti, 2015; Severson & Carlson, 2010). For example, when DVAs pay compliments on appearance, even young children do not hesitate to express their doubts about the authenticity and sincerity of the compliment rather than *pretending* it is a genuine comment about their appearance.

One final consideration is whether DVAs constitute *marginal* entities within children’s emerging ontologies; i.e. entities without a clear categorical belonging. In support of the notion of marginality, children’s perception of DVAs show both ontological firmness and openness: there seems to be a strong awareness of DVAs’ artefactual nature as humanly made and programmable entities, yet the children interpret various aspects of DVAs’ response behaviours as genuine acts of human interference. Moreover, the uniqueness of human language plays a dominant role in children’s understanding about humanness in general and HI in particular, suggesting that certain entities, such as DVAs, might be able to prove their ontological marginality to children through their mastery of human language.

But some caution is required in interpreting these findings in terms of *true marginality*, because children’s interpretations of DVAs’ response behaviours as being genuinely humanoid or non-humanoid could simply be an artefact of the particular VAIG method used in this inquiry. This caution is warranted to the extent that we cannot assume that children would have interpreted DVAs’ response behaviours in the same way if they had not believed in the possibility of human interference. For instance, if children mistakenly address a DVA with the wrong wake word but the machine still displays a reaction (e.g. interaction with the Google Assistant quoted in 4.4), children’s interpretation presumably differs depending on the context: in normal interactions (i.e. *no* believe in the possibility of human interference), children might interpret the DVAs’ reaction as a random technological error rather than a serious sign of DVAs’ ontological marginality. In contrast, in the context of VAIG, children were conditioned to consider every response of the DVA

as a possible sign of human interference, and as a result this conditioning was able to reveal how children associated such error proneness with human nature rather than machines. Therefore, although VAIG may not capture DVAs' perceived marginality as it would become manifest in children's normal open interactions with DVAs, we argue that the method does allow researchers to reveal how children's subjective (and arguably latent) understanding of what it means to be a human and a machine maps onto their observed peculiarities of DVAs. Another strength of VAIG is that it does not require the researcher to make any prior assumptions about children's understanding of what it means to be a human and a machine, because their understanding is revealed implicitly during activity itself. VAIG also allows for open interactions in which children can enjoy lots of freedom how to interact with the devices, which matches the natural interactions they usually experience with DVAs in their home environments or elsewhere.

It is not yet clear, however, whether this indicates (following the general claim of NOCH) that DVAs could constitute an ontological category in their own right. A conclusive answer is certainly beyond the scope of this exploratory inquiry. But our empirical finding that DVAs seem to have humanoid and non-humanoid interactive capabilities which children are able to relate to their subjective understanding of what it means to be a human and a machine does suggest how these devices might establish themselves on the edge between the living and non-living core domains of children's cognitive belief system.

Some of the youngest children in our sample with limited external or even no prior DVA exposure (i.e. sub-group 2.1) also displayed very different interaction patterns and ontological perceptions compared to the older children. This is in line with previous research arguing that age (Brink, Gray, & Wellman, 2019; Cameron et al., 2017; Druga et al., 2017; Kim et al., 2019) or the degree and length of technological exposure (Bernstein & Crowley, 2008; Kory-Westlund & Breazeal, 2019)

are important factors when it comes to children's interactions with, and perceptions of, intelligently behaving machines.

Conclusions, Limitations & Future Research

In light of the introductory rationale regarding the significantly changing real-world context in which today's children grow up due to the increasing presence of DVAs in their environments, this paper argued that any attempt to explore or understand young children's interactions with, and perception of, DVAs should be based on the theoretical grounds of an ontological framework that considers children's genuine understanding of means to be human and what it means to be a machine. Based on focus groups and a gamified data collection design, this small-scale empirical inquiry explored primary school children's ($n=27$, age range: 6 to 10 years, average age: 8.6 years) open interactions with DVAs, how DVAs were embedded in children's general ontological belief system, and how children interpreted certain aspects of DVAs' interactive capabilities as being genuinely humanoid or non-humanoid.

The main findings suggest that children's interactions with DVAs might be more an end in itself than a means to an end, because at times children experience strong notions of engaging excitement whilst conversing with DVAs, without making use of any concrete digital functionalities and services. Secondly, within children's ontological belief system, machines remain the servants of their makers, further implying the superiority of the latter in terms of morality as well as intelligence – although children do acknowledge certain strengths of MI, such as accuracy and speed. Thirdly, the empirical finding that DVAs seem to have observable humanoid and non-humanoid interactive capabilities which children are able to relate to their subjective understanding of what it means to be a human and a machine suggests how these devices might establish themselves on the edge between the living and non-living core domains of children's ontological belief system.

However, these preliminary findings come with an array of limitations, and the theoretical question raised in this paper on children's ontological understanding of digital voice assistants can certainly not be answered conclusively based on such a small-scale, exploratory, qualitative inquiry. In particular, the questions of causality, and whether children's interactions with DVAs alter the perceived characteristics of *existing* ontological categories or yield to a *new* one (i.e. NOCH), need further research and alternative designs which allow for causal claims – as do the questions raised through this inquiry about how children's perceptions of DVAs and respective interaction patterns develop with age, or across different genders and parental influences. Unfortunately, our sample lacked heterogeneity and size, therefore not allowing for a deeper and more rigorous investigation of developmental aspects or other factors that could be relevant from a theoretical perspective (e.g. age, prior DVA exposure, parental backgrounds). Nevertheless, we hope this paper constitutes a fruitful starting point for future research on children's open interactions with, and perceptions of DVAs in order to validate, extend, or criticise the theoretical or empirical arguments we have raised here.

Declaration of Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. No external funding sources have been used to carry out this research.

References

- Airenti, G. (2015). The Cognitive Bases of Anthropomorphism: From Relatedness to Empathy. *International Journal of Social Robotics*, 7(1), 117–127. <https://doi.org/10.1007/s12369-014-0263-x>
- Beirl, D., Yuill, N., & Rogers, Y. (2019). Using Voice Assistant Skills in Family Life. *CSCL 2019*

- *13th International Conference on Computer Supported Collaborative Learning*, 96–103.

International Society of the Learning Sciences (ISLS).

Beneteau, E., Richards, O. K., Zhang, M., Kientz, J. A., Yip, J., & Hiniker, A. (2019).

Communication Breakdowns Between Families and Alexa. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–13. ACM.

Bernstein, D., & Crowley, K. (2008). Searching for Signs of Intelligent Life: An Investigation of

Young Children's Beliefs About Robot Intelligence. *The Journal of The Learning Sciences*, 17(2), 225–247. <https://doi.org/10.1080/10508400801986116>

Biele, C., Jaskulska, A., Kopec, W., Kowalski, J., Skorupska, K., & Zdrodowska, A. (2019). How

Might Voice Assistants Raise Our Children? In W. Karwowski & T. Ahram (Eds.), *International Conference on Intelligent Human Systems Integration. IHSI 2019. Advances in Intelligent Systems and Computing* (pp. 162–167). Cham: Springer International Publishing.

Bohn, D. (2019). Amazon Says 100 Million Alexa Devices Have Been Sold — What's Next?

Retrieved October 13, 2019, from The Verge website:
<https://www.theverge.com/2019/1/4/18168565/amazon-alexa-devices-how-many-sold-number-100-million-dave-limp>

Bonfert, M., Spliethöver, M., Arzaroli, R., Lange, M., Hanci, M., Porzel, R., ... Porzel, R. (2018).

If You Ask Nicely: A Digital Assistant Rebuking Impolite Voice Commands. *Proceedings of the 2018 International Conference on Multimodal Interaction (ICMI '18)*, 95–102.
<https://doi.org/10.1145/3242969.3242995>

Braun, V., & Clarke, V. (2006). Using Thematic Analysis in Psychology. *Qualitative Research in*

Psychology, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness Creeps In: Uncanny Valley Feelings

- Are Acquired in Childhood. *Child Development*, 90(4), 1202–1214.
<https://doi.org/10.1111/cdev.12999>
- British Educational Research Association. (2011). *Ethical Guidelines for Educational Research*.
- Cameron, D., Fernando, S., Cowles-Naja, E., Perkins, A., Collins, E., Millings, A., ... Prescott, T. (2017). Children's Age Influences Their Use of Biological and Mechanical Questions Towards a Humanoid. In Y. Gao, S. Fallah, Y. Jin, & C. Lekakou (Eds.), *TAROS'17 -18th Annual Conference Towards Autonomous Robotic Systems* (pp. 290–299).
<https://doi.org/10.1007/978-3-319-64107-2>
- Canalys. (2019). *Canalys: Global smart speaker installed base to top 200 million by end of 2019*. Retrieved from https://www.canalys.com/static/press_release/2019/Canalys-pr20190415Global-smart-speaker-installed-base-to-top-200-million-by-end-of-2019.pdf
- Carey, S., & Gelman, R. (1991). *The Epigenesis of Mind: Essays on Biology and Cognition*. Hillsdale, England, UK: Erlbaum.
- Castro-González, A., Tan, X. Z., Carter, E., & Steinfeld, A. (2018). Social Haptic Interaction between Robots and Children with Disabilities. *ACM/IEEE International Conference on Human-Robot Interaction*, 81–82. <https://doi.org/10.1145/3173386.3176984>
- CUREC. (2019). CUREC Approved Procedures. Retrieved July 1, 2019, from Research Support website:
<https://researchsupport.admin.ox.ac.uk/governance/ethics/resources/ap#collapse397151>
- Druga, S. (2018). *Growing Up With AI - Cognimates: from coding to teaching machines*. Massachusetts Institute of Technology.
- Druga, S., Williams, R., Breazeal, C., & Resnick, M. (2017). “Hey Google is it OK if I eat you?” *Proceedings of the 2017 Conference on Interaction Design and Children - IDC '17*, 595–600.

<https://doi.org/10.1145/3078072.3084330>

- Druga, S., Williams, R., Park, H. W., & Breazeal, C. (2018). How Smart Are the Smart Toys?: Children and Parents' Agent Interaction and Intelligence Attribution. *Proceedings of the 17th ACM Conference on Interaction Design and Children*, 231–240. <https://doi.org/10.1145/3202185.3202741>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Gaudiello, I., Lefort, S., & Zibetti, E. (2015). The ontological and functional status of robots: How firm our representations are? *Computers in Human Behavior*, 50, 259–273. <https://doi.org/10.1016/j.chb.2015.03.060>
- Gelman, S. A. (1988). Children's Expectations Concerning Natural Kind Categories. *Human Development*, 24(1), 28–34. <https://doi.org/10.1159/000273201>
- Gelman, S. A. (2013). Concepts in Development. In P. D. Zelazo (Ed.), *The Oxford Handbook of Developmental Psychology (Vol 1): Body and Mind* (Vol. 1, pp. 542–563). <https://doi.org/10.1093/oxfordhb/9780199958450.013.0019>
- Harari, Y. N. (2018). *21 lessons for the 21st century*. London, England, UK: Jonathan Cape.
- Harnad, S. (1991). Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem. *Minds and Machines*, 1(1), 43–54.
- Jipson, J. L., & Gelman, S. A. (2007). Robots and Rodents: Children's Inferences About Living and Nonliving Kinds. *Child Development*, 78(6), 1675 – 1688. [https://doi.org/10.1016/0045-6535\(89\)90475-X](https://doi.org/10.1016/0045-6535(89)90475-X)

- Jipson, J. L., Gülgöz, S., & Gelman, S. A. (2016). Parent-child conversations regarding the ontological status of a robotic dog. *Cognitive Development*, 39, 21–35.
<https://doi.org/10.1016/j.cogdev.2016.03.001>
- Kahn, P. H. J., Kanda, T., Ishiguro, H., Freier, N. G., Severson, R. L., Gill, B. T., ... Shen, S. (2012). “Robovie, you’ll have to go into the closet now”: Children’s social and moral relationships with a humanoid robot. *Developmental Psychology*, 48(2), 303–314.
<https://doi.org/10.1037/a0027033>
- Kahn, P. H. J., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., ... Gill, B. (2011). The New Ontological Category Hypothesis in Human- Robot Interaction. *HRI’11 - 6th Annual Conference for Basic and Applied Human- Robot Interaction Research*, 159–160. Lausanne, Switzerland.
- Kim, M., Yi, S., & Lee, D. (2019). Between living and nonliving: Young children’s animacy judgments and reasoning about humanoid robots. *PLoS ONE*, 14(6), 1–16.
- Kinsella, B., & Mutchler, A. (2019). *Voice Assistant Consumer Adoption Report*. Retrieved from <https://voicebot.ai/wp-content/uploads/2019/01/voice-assistant-consumer-adoption-report-2018-voicebot.pdf>
- Kory-Westlund, J. M., & Breazeal, C. (2019). Assessing Children’s Perceptions and Acceptance of a Social Robot. *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 38–50. ACM.
- Ku, H., Choi, J. J., Lee, S., Jang, S., & Do, W. (2018). Designing Shelly, a Robot Capable of Assessing and Restraining Children’s Robot Abusing Behaviors. *ACM/IEEE International Conference on Human-Robot Interaction*, 161–162.
<https://doi.org/10.1145/3173386.3176973>

- Lee, K.-F. (2018). *AI superpowers: China, Silicon Valley, and the new world order*. Boston, MA, US: Houghton Mifflin Harcourt.
- Lovato, S. B., Piper, A. M., & Wartella, E. A. (2019). Hey Google, Do Unicorns Exist?: Conversational Agents as a Path to Answers to Children's Questions. *Proceedings of the 18th ACM International Conference on Interaction Design and Children*, 301–313. ACM.
- Lovato, S., & Piper, A. M. (2015). "Siri, is this you?": Understanding Young Children's Interactions with Voice Input Systems. *Proceedings of the 14th International Conference on Interaction Design and Children - IDC '15*, 335–338. <https://doi.org/10.1145/2771839.2771910>
- Müller, K., & Schulte, C. (2018). Are children perceiving robots as supporting or replacing humans? *Proceedings of the 13th Workshop in Primary and Secondary Computing Education*, 1–4. <https://doi.org/10.1145/3265757.3265767>
- Mutchler, A. (2017). Voice Assistant Timeline: A Short History of the Voice Revolution. Retrieved November 5, 2019, from voicebot.ai website: <https://www.voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/>
- Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, 1–12. <https://doi.org/10.1145/3173574.3174214>
- Postbank. (2019). *Postbank Digitalstudie 2019: Untersuchung zur Internet und Mobilnutzung der Bevölkerung*. Bonn, Germany.
- Pradhan, A., Findlater, L., & Lazar, A. (2019). Phantom Friend or Just a Box with Information: Personification and Ontological Categorization of Smart Speaker-based Voice Assistants by

- Older Adults. *Proceedings of the ACM on Human-Computer Interaction*, 1–21. ACM.
- Quantcast. (2019). *Quantcast Audience Insights: Sprachassistentin Alexa spricht vor allem Familien mit Kindern an*. Munich, Bavaria, Germany.
- Reeves, B., & Nass, C. I. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge, England, UK: Cambridge University Press.
- Ross, C. (2019). New voices at patients' bedsides: Amazon, Google, Microsoft, and Apple. Retrieved February 28, 2019, from STAT website: https://www.statnews.com/2019/02/06/voice-assistants-at-bedside-patient-care/?utm_campaign=the_download.unpaid.engagement&utm_source=hs_email&utm_medium=email&utm_content=69643770&_hsenc=p2ANqtz--nOHcxWiobOCFlnUebi0I5Zc_2HUuv1GHjLPLxaPehTUojtzL-mYVdu0NjNz
- Saygin, A. P., Cicekli, I., & Akman, V. (2001). Turing Test: 50 Years Later. *Minds and Machines*, 10, 463–518.
- Scaife, M., & Van Duuren, M. (1995). Do computers have brains? What children believe about intelligent artifacts. *British Journal of Developmental Psychology*, 13(4), 367–377. <https://doi.org/10.1111/j.2044-835X.1995.tb00686.x>
- Schwab, K. (2017). *The Fourth Industrial Revolution*. Geneva, Switzerland: World Economic Forum.
- Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018). Hey Alexa, What's Up?: A mixed-methods studies of in-home conversational agent usage. *Proceedings of the 2018 Designing Interactive Systems Conference*, 857–868. <https://doi.org/10.1145/3196709.3196772>
- Seibt, J. (2017). Towards an Ontology of Simulated Social Interaction: Varieties of the “As If” for

- Robots and Humans. In R. Hakli & J. Seibt (Eds.), *Sociality and Normativity for Robots: Philosophical Inquiries into Human-Robot Interactions* (pp. 11–39).
<https://doi.org/10.1007/978-3-319-53133-5>
- Severson, R. L., & Carlson, S. M. (2010). Behaving as or behaving as if? Children’s conceptions of personified robots and the emergence of a new ontological category. *Neural Networks*, 23(8–9), 1099–1103. <https://doi.org/10.1016/j.neunet.2010.08.014>
- Small, C. E., Nigrin, D., Churchwell, K., & Brownstein, J. (2018). What Will Health Care Look Like Once Smart Speakers Are Everywhere? Retrieved February 28, 2019, from Harvard Business Review website: <https://hbr.org/2018/03/what-will-health-care-look-like-once-smart-speakers-are-everywhere>
- Tanaka, F., Cicourel, A., & Movellan, J. R. (2007). Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences*, 104(46), 17954–17958. <https://doi.org/10.1073/pnas.0707769104>
- Trippas, J. R., Spina, D., Scholer, F., Awadallah, A. H., Bailey, P., Bennett, P. N., ... Salim, F. D. (2019). Learning About Work Tasks to Inform Intelligent Assistant Design. *Proceedings of the 2019 Conference on Human Information Interaction and Retrieval*, 5–14. ACM.
- Turing, A. (1950). Computing machinery and intelligence. *Mind*, 49, 433–460.
- Turkle, S. (2005). *The Second Self: Computers and the Human Spirit*. Retrieved from https://monoskop.org/images/5/55/Turkle_Sherry_The_Second_Self_Computers_and_the_Human_Spirit_20th_ed.pdf
- Vollmer, A. L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science Robotics*, 3(21). <https://doi.org/10.1126/scirobotics.aat7111>

- Watt, S. (1996). Naive Psychology and the Inverted Turing Test. *Psychology*, 7(14).
- Waytz, A., Cacioppo, J., & Epley, N. (2014). Who Sees Human? The Stability and Importance of Individual Differences in Anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232. <https://doi.org/10.1177/1745691610369336>.Who
- Wellman, H. M., & Gelman, S. A. (1992). Cognitive Development: Foundational Theories of Core Domains. *Annual Review Psychology*, 43, 337–375.
- Wood, L. J., Robins, B., Lakatos, G., Syrdal, D. S., Zaraki, A., & Dautenhahn, K. (2018). Piloting Scenarios for Children with Autism to Learn About Visual Perspective Taking. In M. Giuliani, T. Assaf, & M. E. Giannaccini (Eds.), *Proceedings of the 19th Annual Conference Towards Autonomous Robotic Systems - TAROS 2018* (pp. 260–270). <https://doi.org/10.1007/978-3-319-96728-8>
- Złotowski, J., Proudfoot, D., Yogeewaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human – Robot Interaction. *International Journal of Social Robotics*, 7, 347–360. <https://doi.org/10.1007/s12369-014-0267-6>

Appendix

Appendix I. Sub-Group composition during VAIG

Primary School 2					
Sub-Group 1.1			Sub-Group 1.2		
Child	Gender/Age	DVA Exposure	Child	Gender/Age	DVA Exposure
James	M/8	Yes (domestic)	Lukas	M/9	Yes (domestic)
Megan	F/8	Yes (external)	Emily	F/10	Yes (external)
Leon	M/8	Yes (external)	Jeff	M/9	Yes (external)
Laura	F/8	Yes (external)	William	M/10	No (-)
Paul	M/9	No (-)	Natalie	F/10	Yes (external)
Marianne	F/8	Yes (external)	Manuela	F/9	Yes (domestic)

Primary School 2					
Sub-Group 2.1			Sub-Group 2.2		
Child	Gender/Age	DVA Exposure	Child	Gender/Age	DVA Exposure
Freya	F/9	No (-)	Jordan	M/9	Yes (external)
Alice	F/6	No (-)	Kate	F/9	Yes (external)
Elisa	F/6	Yes (external)	Jonas	M/9	Yes (domestic)
Clara	F/6	Yes (domestic)	Max	M/8	No (-)
Steven	M/7	No (-)	Stephanie	F/8	Yes (external)

Notes. Table provides details of the sub-Group composition during VAIG for both primary schools. For the sake of full anonymity, all names are pseudonyms with no similarity to the real names of individuals. DVA-exposure refers to the presence of DVAs at home (domestic) or within the closer circle of family in friends (external) according to the parental survey.