

JUST LIKE ME
BUT NOT
EXACTLY

AI, ANTHROPOMORPHISM
& THE HUMAN-
TECHNOLOGY GAP

IULIA IONESCU
ROYAL COLLEGE OF ART
APRIL 2022

JUST LIKE ME,
BUT NOT EXACTLY:
AI, ANTHROPOMORPHISM &
THE HUMAN-TECHNOLOGY GAP

IULIA A. IONESCU

A THESIS SUBMITTED IN FULFILMENT
OF THE REQUIREMENTS OF THE
ROYAL COLLEGE OF ART FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY.

01/04/2022

ABSTRACT

While artificial intelligence (AI) design continues to aspire towards romanticised visions of intelligence and human-likeness, the increasing dichotomy between its technological and social facets has given rise to tangible evidence of surveillance, inequality and discrimination. This evidence emphasises a pressing need for systems more aligned with human expectations and values, but computationally, it remains extremely difficult to embed concepts as fundamentally fluid and situational as value and meaning. Such difficulty has led present-day research to formalise these intricate principles of human behaviour, in all their complexity, through predictive machine learning models built on data extracted from how people behave not in relationship to AI but in the absence of it.

This research proposes that the inseparability of the AI social artefact from the Human entails, foremost, an understanding of the interaction between concept (humanlike design) and percept (anthropomorphism). Consequently, the effect of these variables on each other is explored in two stages: first by reviewing whether our mental models of humanlike AI are built on pre-existing knowledge and schemas of other people (Just Like Me), and second, by probing their homogeneity with respect to their corresponding human counterparts (But Not Exactly). Building on the results of these initial investigations, this study demonstrates that the locus of the interaction, in practice, cannot be neatly extracted from manipulating the correlation of these variables, but rather by looking at how it is continuously constituted in the dynamic between the two. With combinations of effects born in the fleeting concept-perception interaction, it is concluded that very few behavioural patterns and mental models can truly be universalised, and for the most part, one context cannot be neatly applied to the next.

To address this epistemological limitation, the present study proposes a novel methodological approach – taking influence from cognitive dissonance theory – that requires the designer/researcher to locate themselves in the complexity of this socio-technical world and to reflect on how their interventions impact that which they are observing. In addition to providing a visual method of investigating the interaction, the proposed approach is ultimately intended as a design strategy for addressing the nonlinear and often chaotic dynamics of an emergent field of study: the Human-AI.

LIST OF FIGURES

All figures included in this thesis are personal work unless stated otherwise.

1 – Research Journey Matrix	22
2 – Research Framework Diagram	31
3 – Canepa Motorsport (2020), 1959 Austin-Healey Sprite [Accessed 27 February 2022]. Available from: flickr.com	40
4 – Canepa Motorsport (2022), 20 Ford GT 6634 [Accessed 27 February 2022]. Available from: flickr.com	40
5 – Pieters, Geert (2020) Jean Paul Gaultier Classique Bottle [Accessed 27 February 2022]. Available from: unsplash.com	41
6 – Spaa, Joppe (2020) Azzaro Chrome Eau de Parfum for Men [Accessed 27 February 2022]. Available from: unsplash.com	41
7 – Anthropomorphic Predictions	51
8 – Anthropomorphism does not statistically vary over time	58
9 – One-way ANOVA with anthropomorphism and understanding chart	58
10 – One-way ANOVA with anthropomorphism and user satisfaction chart	59
11 – One-way ANOVA with anthropomorphism and chat feature chart	59
12 – Predictions vs Results	63
13 – Word Cloud Analysis	67
14 – Separating Affordances from the Perceptual Information that specifies affordances [adopted from (McGrenere,2000)]	77
15 – LED Feedback Panel Design	82
16 – Physical Device Design	83
17 – Experiment Design (Procedure)	85
18 – Experiment Data Analysis	87
19 – Physical Design (Final Object)	99
20 – Physical Design Prototype / Meeting Echo's Module Requirments	100
21 – Device Exploded View	101
22 – Results Literal Condition	104

23 - Results Robotic Condition	105
24 - Overlapping Agent Knowledge Estimates	108
25 - Cognitive Dissonance Framework	121
25 - Legend Generative Parameters	130
27 - Legend Structural Annotations	131
28 - Siri Case Study Topography	134
29 - Siri Case Study Detail	135
30 - Google Home Case Study Topography	138
31 - Google Home Case Study Detail	139
32 - Alexa Case Study Topography	140
33 - Alexa Case Study Detail	141

ACKNOWLEDGEMENTS

I would firstly like to thank Microsoft Research Cambridge for funding this research, and my two supervisors, Prof. Ashley Hall and Richard Banks, for their continuing guidance and mentorship.

I'd also like to thank Jack O'Leary McNeice, whose pedantic observations about my use of the word 'respectively' have considerably improved the legibility of this thesis.

Finally, I would like to thank my mom for her infinite and unwavering support. I love you, and I know you would have been very proud.

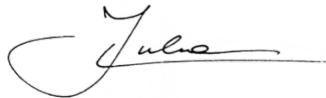
As for my dad, had he known I would start a PhD, he would have told me not to.

AUTHOR'S DECLARATION

During the period of registered study in which this thesis was prepared the author has not been registered for any other academic award or qualification. The material included in this thesis has not been submitted wholly or in part for any academic award or qualification other than that for which it is now submitted.

Signature: Iulia A. Ionescu

Date: 01/04/2022

A handwritten signature in black ink, appearing to read 'Iulia', with a long horizontal flourish extending to the right.

ABSTRACT	1
LIST OF FIGURES	2
ACKNOWLEDGEMENTS	4
AUTHOR'S DECLARATION	5

00. Introduction **8**

CONTEXT	10
THE KNOWLEDGE GAP	13
SIGNIFICANCE	19
STRUCTURE	21
METHODOLOGICAL UNDERPINNING	24

01. The Human **32**

PROLOGUE	33
IOT <> AI	34
AI <> ANTHROPOMORPHISM	36
ANTHROPOMORPHISM <> DESIGN	38
THE HUMAN <> ANTHROPOMORPHISM	46
A. ELICIT AGENT KNOWLEDGE	47
B. EFFECTANCE MOTIVATION	48
C. SOCIALITY MOTIVATION	50
[EXP01] THE LOOPHOLE	52
A. INTRO	52
B. METHOD	53
C. FINDINGS	61
DISCUSSION	68

02. The AI **72**

AFFORDANCES <> ANTHROPOMORPHISM	74
[EXP02] JUST LIKE ME	80
A. METHOD	80
B. FINDINGS	89
[EXP03]...BUT NOT EXACTLY	93
A. METHOD	94
B. FINDINGS	107
DISCUSSION	110

03. The Human-AI **114**

HUMAN-AI <> COGNITIVE DISSONANCE	116
A. THEORETICAL ENTITIES	118
[EXP04] INTERSPACE	123
B. EXISTING WORK	125
C. METHOD	126
D. FINDINGS	133
E. DESIGNING ONWARDS	142

04. Conclusions **144**

OVERVIEW	145
CONTRIBUTION TO KNOWLEDGE	146
A. THEORY	146
B. METHODOLOGY + METHOD	149
C. DESIGN PRACTICE	151
FUTURE RESEARCH	153

List of References **156**

05. Appendices **178**

A. GLOSSARY OF TERMS	179
B. PHD TIMELINE	183
C. VIRTUAL ASSISTANTS SURVEY	184
D. VA SURVEY SUMMARY	188
E. HUMAN NATURE AND HUMAN UNIQUENESS SCALE	238
F. 'JUST LIKE ME' – ACTIVITY	240
G. 'JUST LIKE ME' – DEVICE DESIGN	244
H. 'BUT NOT EXACTLY...' SURVEY	248
I. 'BUT NOT EXACTLY...' SURVEY SUMMARY	249
J. 'BUT NOT EXACTLY...' DEVICE DESIGN	252
K. INTERSPACE: CONVERSATIONAL ANALYSIS	256
L. INTERSPACE: DATA ANALYSIS	261
M. INTERSPACE: CONCEPT STUDIES	264
N. CONSENT FORM SAMPLE	268
O. SOURCE CODE	270

00
INTRO

The idea of AI has been around for decades, emerging from a narrative tacitly laden with romanticised visions of intelligence and automation. While these ideals remain, in part, laid out only for the purpose of thought experimentation, substantial leaps have brought AI into the realm of observable and testable reality, with tangible consequences for the way we live and hope to live.

With the advance of technology, AI development has continued to converge towards one of its foundational goals of making machines do humanly things. As a result, automation was not enough to play the ‘imitation game’, and the deep desire to create intelligent machines has become virtually synonymous with a desire to also integrate such machines into our social milieu (Proudfoot, 2011). This aspiration has meant, in part, identifying and testing several fundamentally human concepts, with machines that have now become able to dependably fill many of the roles previously considered exclusively human: doctors, financial analysts, judges, assistants, chauffeurs (Sloane and Moss, 2019).

While AI’s profound impact on our lives has been brought to the forefront by two substantial technological developments – big data and computing power – the enthusiasm fuelling its adoption has more likely been sustained by its initial foresight: the sentimental aspiration towards human-likeness (Hayes and Ford, 1995; Levesque, 2017). Consequently, anthropomorphism (the act of attributing humanlike qualities to nonhuman agents) has not merely managed to find its way into designers’ repertoires of problem-solving tools but has been almost entirely assimilated at the core of this socio-technological phenomenon that continues to redefine the very fabric of our social reality.

However, the focus on anthropomorphism has given rise to a whole new set of limitations, with the equivocal nature of this metaphor only serves to further obfuscate our expectations of an already mystified technological system. The dissonance between our anthropomorphically-inclined expectations and AI’s technological reality has led to a tendency to disconnect AI’s technical and social facets in practice.

Although this research concentrates on exploring the nature of an emerging dynamic between the Human and the social aspects of AI – with anthropomorphism forming a testable point of convergence between the social and technological consequences of factions of AI – it is first essential to acknowledge not only the broader disconnect between AI’s technical design and its social effects in practice, but also the relative scarcity of transdisciplinary approaches in research.

CONTEXT

With the broadening prospect of interacting with social machines wide on the horizon, the space between AI's technological and social facets has continued to expand, generating a gradually increasing underexplored gap (Hayes and Ford, 1995). For the most part, AI, with its organised 0s and 1s, continues to conjure the image of a merely technological tool that, much like its logic, can be switched off and on inconsequentially (Dahlin, 2021). However, this prevalent construct remains somewhat misleading since many of these AI 'tools' have, in fact, ended up entirely changing the premises upon which our social practices are based (Callon, 2005). The stock market is a salient example of this: the introduction of AI systems has entirely reshaped the conditions of the market into a practice that would have not even existed in its absence (Rundle, 2019). Undoubtedly, changes are inherent to the adoption of innovative tools and technologies, however, some AI systems are unique in the sense that they have become fully-fledged actors with undeniable influence over our human behavioural practices. With humans trading alongside AI systems, the role of a financial analyst has evolved to anticipate not only the psychology of human fears and demands but also how other AI systems would behave in response to these dynamics (Lenglet, 2011). Consequently, AI systems, irrespective of how 'social' their design conceptualisation is, tend to intervene with the conditions of our social practices, consistently leading to humans shifting their identities, behaviour and reasoning in response. As a result, the algorithms are rendered inherently 'social' in the process.

Even in the face of the – now broad – acknowledgement of AI's social impact, its interpretation as being first and foremost a technological object entails that often, all other considerations come in at a very distant second (Sloane and Moss, 2019). This has been particularly evident with systems that we now know perpetuate harmful beliefs and exacerbate societal biases. From governmental automated decision-making systems that increase stigma through prejudiced profiling based on class, gender or race (Sztandar-Sztanderska and Zielenska, 2018) to demonstrably higher error rates for pedestrians with darker skin tones in object detection algorithms (Wilson, Hoffman and Morgenstern, 2019), or the significant gender stereotypes in word embedding (a popular framework of representing text data) (Bolukbasi *et al.*, 2016), algorithmic failures can be seen as frequently being the result of misconstruing or overlooking AI's social impact.

Moreover, detaching AI's technological embodiment from its social manifestation may happen at any point in its timeline, from concept generation to technical design and implementation, making it all the more challenging to recognise and address the space between its technological and social facets.

The significance of this gap has received increased attention in recent years. For instance, Amazon's 2014 talent acquisition specialist, AMZN.O – a machine learning tool programmed to review job applicants' resumes – morphed from what was initially considered a holy grail of automation into an agent of discrimination (Dastin, 2018). It took a year for the company to discover that their tool was not rating candidates in a gender-neutral way as, due to benchmarking their training dataset against the successful resumes submitted to the company over a 10-year period, the training data for the model echoed the male dominance across the tech industry, which in turn led AMZN.O to teach itself that candidates who were less-male were also less-preferable for technical jobs. Amazon later changed their algorithms to account for this bias, but the damage was far greater than this shortcoming that nevertheless took a year to be uncovered. The possibility of these machines devising other discriminatory processes that may go unnoticed gave birth to a lasting fear. A very similar weakness crept in Poland's Public Employment Services AI system, designed to automate the process of directing unemployed citizens towards the programmes they were most eligible for (Sztandar-Sztanderska and Zielenska, 2018). In this instance, however, individual profiling was based on data collected by case counsellors who obscured the purpose of its use from participants (Sztandar-Sztanderska and Zielenska, 2020). As a result, the system ended up unfairly profiling individuals into good or bad investments based on a logic that defied scrutiny, which in turn led the Human Rights Commissioner to rule the AI as unjust, eventually banning it from public use.

These failures, however, are not always the result of poor technical design. In 2016, Microsoft launched Tay (short for Thinking About You), a chatbot with the premise of advancing our understanding of both natural language processing (NLP) and anthropomorphic conversations (Lee, 2016). Its downfall, this time, came from the context in which it was set to operate: the chatbot was launched on Twitter, a platform that, its developers confidently anticipated, would contribute to Tay's conversational skills. However, Tay was programmed to mirror natural conversational practices and to hold better dialogues by learning from interactions with Twitter users, which it did very well. In fact, it did so well that it prompted Microsoft

to pull the plug on the project in less than 24 hours after its launch because it only took that long for Twitter users to spot its vulnerabilities and to ‘hack’ its learning algorithms with racist, misogynistic and anti-Semitic tweets, turning the chatbot into a mouthpiece for a chilling ideology.

While these examples may appear to indicate the need for a more rigorous analysis of datasets and machine learning approaches, in fact, AI’s social component has a far broader reach. Even systems that seem to prodigiously tackle technical problems – such as in the case of computer vision tools capable of outperforming doctors in interpreting screening results – usually end up being intolerable in practice (Topol, 2019). No AI system is completely removed from humans, and its functionality and acceptance in a fundamentally social environment depends greatly on it being tolerated by humans. The typical lack of scrutiny and transparency when it comes to AI further impacts the delicate trust involved in Human-AI relations (Watson *et al.*, 2019), adding yet another layer of social considerations that usually slip through the cracks of these disjointed facets of study. For the practice of designing AI systems to overcome matters such as these, the inseparability of the socio-psychological dimensions from technical concerns needs to be better reflected in both research and practice.

However, this gap between technology and society has, arguably, always been somewhat contradictory since all software systems are already social systems, but the AI systems we will be examining in this research are those that go beyond this often-covert interaction.¹ Such systems are those purposefully designed to act as social agents, and those that are planned as social objects yet still largely designed as technical ones² – drawing explicit attention to the gap in AI development and the scarcity of intradisciplinary research and practice. As argued by Dahlin (2021), it may simply not be enough for engineers to adopt

1 AI systems are often portrayed through the lens of their statistical probabilities and performance metrics to evoke trust. The act of quantification carries with it the covert implication of a scientific credibility devoid of human interpolation, interpretation or error.

2 Therefore, to reflect this context when speaking about AI, this research will exclude machine learning systems that are not set to operate directly with humans in a highly social setting, such as (but not limited to): fraud and spam detection, generative models or certain recommendation algorithms. For this reason, the word AI will be used synonymously with ‘social technology/machine’ or ‘agent’ as a reminder of the social backdrop of this exploration.

a social science perspective or for social scientists to learn machine learning skills. This is an opportunity to explore how the practice of designing the smart technology of the future might look like were the knowledge and approaches of these disparate but equally relevant disciplines to be interwoven with its development.

With a socio-psychological perspective having tacit but significant implications for AI's technical design, this research aims to explore this space between the Human and the AI, in a practice that draws inspiration from the various contributing fields and tries to explore the potential for a methodological merger and a common language of analysis.

THE KNOWLEDGE GAP

AI has long been the engine of advanced STEM research, with a lot of the discussions surrounding it being held in largely technical circles and fixated on barriers to advances or subtechnological applications (Zhang and Aslan, 2021). However, the broader appeal and advantages of AI were simply too tantalising to exist within this academic vacuum, and it hasn't taken long for intelligent machines to start saturating a broad range of social domains, ranging from manufacturing to education, from healthcare to retail and from transport to law. Early on in its development, it was assumed that AI would revolutionise simple repetitive tasks involving low-level decision-making, however, its precipitous growth in sophistication in a very short amount of time has seen a shift in our understanding of what constitutes AI and what it entails.

Consequently, over past decades AI stood for many things, from science to artefacts and from processes to phenomena, with research on AI closely mirroring these meanings. AI's early definition as 'the science and engineering of making intelligent machines' (McCarthy, 2007, p. 1) dominated much of the 1980s and 1990s research – research that unsurprisingly saw AI as a system of scientific knowledge concerned with making machines do humanly things (McCarthy *et al.*, 1955). Later on, thanks to advances in machine learning (ML) and deep learning (DL), two techniques on which AI development had become so reliant, its definition shifted to become synonymous with these two terms. AI research, yet again, moved to gravitate towards branches of computer science more interlinked with statistics and mathematics, neuroscience, linguistics

and philosophy (Liu, 2021). More recently, with the mass commercialisation of AI, its meaning has entered the common lexicon via a series of applications and subtechnologies such as the Internet of Things, home assistants, chatbots, big data and so on, which, once again, led to AI research being more concerned with its development as a distinctive socio-cultural phenomenon (Schwartz, 1989). The shift towards our current conception of AI has significantly broadened the range of disciplines interested in critically examining its role in society and led to the proposal of models that are more aligned with human expectations and values.

The extent of literature on this topic not only exceeds the capacity of this research, but any attempt at comprehensiveness here would inhibit the methodological strategy (Grounded Theory), discussed later in this chapter. Therefore, rather than offering a comprehensive survey of the work in this field (which spans across seven decades), in the following few pages, I focused my attention on the work dealing with interactions between humans and AI in social settings. The following framing of the points of intersection between AI's technological and social considerations should, therefore, not be seen as a comprehensive summary of the research into AI, but rather as a consultation with the pertinent literature aimed at exposing potential gaps therein and inspiring the ensuing research questions.

Yet this endeavour proved to be challenging as much of the literature currently purporting to focus on the 'sociology of AI' seems particularly dispersed and dated. Zheng Liu points out in her research on the sociological perspectives of artificial intelligence that "Searching for 'artificial intelligence; sociology' on Google Scholar (GS), [...] can return a remarkably heterogeneous body of results. Some studies analyse the implications of AI for social processes and relations. Others discuss the use of AI techniques in sociological research. Still others appear in the search results because they happen to list 'artificial intelligence' and 'sociology' as keywords, and yet do not analyse 'artificial intelligence' in sociological terms. Moreover, in a GS search conducted on 20 May 2020, 80% of the studies shown on the first results page (results sorted by relevance) were published between 1980 and 2000, and all before 2003. In comparison, 70% of the most relevant results from a search for 'artificial intelligence; economics' was research published in the past decade." (Liu, 2021, p. 2)

As a result, navigating the literature related to AI's social considerations can be challenging without alluding to a specific AI technique or application, such as 'Internet of Things (IoT)', 'robots', 'self-driving cars', 'natural language processing' and so on, which defies the purpose of examining AI as an experience rather

than a tangible object or service. However, this limitation helps to underline the two principal strands in which AI's social and technical facets intersect: AI as an artefact and AI as a phenomenon.

A. AI AS AN ARTEFACT:

Most of the sociological discussions surrounding AI in the 1980s and up until the 2010s saw AI as tangible systems, techniques or products and focused on the social ramifications of its implementation. From very early on, we find concerns around the subjective knowledge embedded in building these systems, such as Forsythe (1993a, 1993b) work on how AI systems embody and reproduce their developers' biases and cultural values and assumptions. Much like in Bloomfield's (Bloomfield, 2018) early work on how AI research is conditioned by subjective cultural and social factors that hide behind a veneer of objectivity, Forsythe (1993a, 1993b) argues that AI artefacts not only reproduce human biases but also bestow them with some form of scientific credibility. This, of course, is not news today, but given that her research was published nearly three decades ago, one is inclined to question why are we still grappling with these issues nearly 30 years later.

In the same vein, we also find a plethora of research that views AI's alleged humanlike abilities with scepticism. Work in this category may draw our attention to AI's social deficit and to its inability to 'see' and 'understand' real context, core human abilities that have proven challenging to replicate. For example, Collins (1992, 2018) argues that, while intelligent machines demonstrate impressive abilities for pattern recognition, they remain mechanical tasks devoid of real understanding. Similarly, Suchman (1987, 2006, 2011) goes to great lengths to demonstrate that robots can only behave in humanlike fashion on the grounds that they have been prescribed with programs that tell them to do so and, that in the absence of an ability to modify their algorithmic logic in response to their context, they will in fact never be able to truly reproduce human behaviour.

Another branch of research in this field focuses on AI's various applications as objects of study, from workforce automation and its repercussions (Frey and Osborne, 2013; Spencer, 2016; Boyd and Holton, 2017; Charles M. A. and Aleksandr V., 2020) to research on self-driving cars and their 'social learning' deficit (Stilgoe, 2018). In this category, we find concerns with various applications needing to be

more technically and socially fit. Research of this kind emphasises the role of social factors in both analysing AI as a social artefact and also in informing its technical makings early on in the design and experimentations phases (Marres, 2020a, 2020b). Similarly, in the context of IoT and smart homes, research spans across the wide range of AI subtechnologies and their impact on lived human experiences. From Fantini's work on challenging the oversimplification of complex, lived, human experiences and proposing a slowing down approach to designing IoT objects (Fantini van Ditmar, 2016), to grappling with personal assistants' perceived autonomy and its consequences on human behaviour, research of this kind challenges our approach to designing AI-powered services or products in light of their social repercussions.

Other researchers in this field examine the rise of privacy concerns and automated surveillance from a socio-psychological perspective, with a particular focus on how some of these technological depictions, as idealised portrayals of domestic service, make "contending with issues such as surveillance and digital labour increasingly difficult" (Phan, 2019). Similarly, another line of research draws attention to the difference between automation and autonomy and a general concern regarding the extent to which we are designing automated systems to appear autonomous. Underlying these concerns is the clear distinction between AI being able to make decisions in a humanlike manner and simply acting 'as if' they were able to make these decisions as moral agents (Waytz *et al.*, 2010).

In all these studies, we notice that with AI's increasing consumption, a primary research interest revolves around how AI's development and portrayal is generating issues across a diverse range of fields of study. Ranging from the dangers of automation to discussions around autonomy and decision-making, the common theme that unifies these studies can be said to be their analytical perspective focused on AI's consequences on society brought upon by its implementation as an artefact. However, much of this style of post-analysis remains a dangerous simplification of how our social world functions – a conclusion that itself resulted from the very same process of post-analysis employed experimentally in the first two chapters of this research. Seeing individuals as the locus of norms and values inevitably leads to the assumption that behaviours shift in a predictable and direct response to technological interventions, which they do not (at least not entirely, as will be demonstrated later in this research).

B. AI AS A PHENOMENON:

In contrast to the previous theme, in which the research discussed has been concerned with more tangible approaches to techniques, products or services employing AI and the social ramifications thereof, research under the banner of 'AI as a phenomenon' is less concerned with AI as an artefact. Instead, it takes AI as an object of study and explores its interaction with the wider social, cultural, political and economic contexts in which it operates and by which it is affected. In this category, we find broader discussions surrounding human metaphors, anthropomorphism, and the cultural constructions of AI, such as discussions of how different cultures leverage different resources to trigger wider social interest and acceptance of AI.

One example of this is Šabanović (2014), who has investigated how Japan's robot culture leverages popular leitmotifs and cherished cultural values in their design of AI-powered robotic interventions, cunningly eliciting cultural consent through appeal to social norms. Similarly, Phan (2019) looks at the depiction of digital assistants as a cultural representation, which in many ways works to reproduce the idealised relationship between servant and master in nineteenth- and twentieth-century American homes. In the same vein, Spigel (2001) defines this romanticised vision of AI as a method of advertising domestic technologies as "futuristic pastoral fantasies": in the case of Amazon Echo, it is "a white, middle-class nuclear family lifestyle", which sentimentalises traditional domestic ideals (Phan, 2019, p. 5).

Still, while these investigations demonstrate the significance of cultural framing in AI development as a necessary consideration in designing AI products and services, they also shed light on how AI functions as a regulator of social values, influencing or reinforcing stereotypes and ideals through cultural framing or de-framing (Liu, 2021). This idea influenced one of the experiments carried out in the course of this research, an experiment aimed at uncovering the mechanism of stereotype formation through cultural framing and design's impact on mental model formation (Chapter II).

It should be noted that AI's influence on human cognition is not a novel research topic and many thinkers, such as Turkle (2005), have explored the effects of computers on sociality since the conception of this research field. In her work, Turkle found that not only do we tend to describe AI using human terms and

analogies, but that the use of human features and terms in association with AI and computers usually leads to the opposite of the intended effect – describing human processes in terms of computer analogies. As a result, she argued that the proliferation of computers and AI will have colonising effects on society, leading to humans starting to think and behave in relationship to each other in computer terms and processes (see Levesque, 2017). This, in turn, means that AI is not simply colonising, but by describing human processes in terms of AI metaphors, it leads people to alienate themselves from their socio-economic conditions, dehumanising them in the process.

Studies such as these, which draw upon perspectives from media and cultural studies, philosophy, sociology and critical design, indicate that AI is not merely an artefact but also a cultural construction with vast repercussions for society's functioning that are not always easy to identify and address. The very occurrence of this type of analysis, of AI as a phenomenon, reflects the accelerating social and "cultural embeddedness" (Liu, 2021) of social algorithms. Despite this, perhaps due to the more conceptual nature of treating AI as a social actor, this shifting perspective often remains relatively underexplored in practice.

Finally, these two emerging typologies in research have proven an effective tool for understanding how AI has been conceptualised in sociological terms, but eventually there may be many other ways in which AI remakes social realities, which we may not have the capacity to fully imagine today. There is no real reason why AI cannot be examined from very different analytical perspectives (Cave, Dihal and Dillon, 2020), and in line with this idea, in this research, I propose a different mode of conceptualisation, at the intersection of its artefact embodiment and its socio-cultural perception – examining it as a mode of social interaction.

As I will continue exploring in this thesis, this conceptualisation of AI as both concept and percept has been obfuscated by the fact that AI's social considerations in research and the practice of designing AI remain greatly out-of-sync. This, as should become clear later, is due to the scarcity of methodologies suited to consider the perception and the conceptualisation synchronously. Accordingly, I will later outline my use of practice-based design research methods in better exploring this new AI conceptualisation, and discuss how such methods may be better suited to addressing this gap.

SIGNIFICANCE

There is still a wide gulf separating the research and methods that social sciences have honed when researching humans interacting with AI and the computational strategies that still guide much of AI design. This gulf is, in part, the effect of very different vocabularies, journals and methods, but it is also the result of pursuing very different questions that inevitably filter out the intricacies of the interaction between the artefact and the phenomenon.

Consider, for example, the ‘moral machine dilemma’ (Awad *et al.*, 2018) and self-driving cars, a context in which equipping AI with basic ethical principles is indispensable to rendering its decisions socially acceptable, particularly in a crash scenario. However, this acknowledgement that machines should act in agreement with human values, while common-sensical from a design or socio-psychological perspective, becomes almost impossible to translate into 0s and 1s, and any attempt to do so dangerously oversimplifies something fundamentally fluid and circumstantial. The issue, however, is not technology’s limitation in encoding fluid human values into machines. It is also not sociology, anthropology or psychology’s limitation in laying down the principles of complex human behaviour and social truths (Sloane and Moss, 2019). The real issue lies in the view that these effects can be extracted from how humans behave in contrived situations, external to any lived experiences or social contexts. Culture, values and meanings are continuously being constituted through the ways people interact not only with each other but also with the AI and, as such, are subject to the circumstances and details of such interactions. Much like in the practice of stock-trading, in which human mental models shift to account for the AI, when it comes to autonomous vehicles, our ethical principles might fluctuate to accommodate for yet unidentifiable factors.

This schism, then, encourages a practice that assumes the premises of the Human-AI phenomenon is inherently stable. To illustrate this, we need not look further than how we are training machine learning models on data extracted from how people behave not *in relationship to AI*, but *in the absence of it*. That is, the *imagined* humans for which we are building these systems are very different from the real-life meaning-constructing humans who change their behaviour and reasoning in response to the complex context of an interaction. This process fundamentally challenges the relevance of the training datasets.

However, ethical and moral principles are only two of many amorphous socio-psychological considerations that cannot be separated from the context of the Human-AI interaction. Very few behavioural patterns and mental models can truly be universalised, and for the most part, one context cannot be neatly applied to another. There are combinations of effects that cannot happen outside of the interaction, which is why exploring the space between these two facets cannot merely be calculated or pre-designed. In order to grapple with the vastness of this gap, we first need to acknowledge that the interaction is not stable but subject to change. This may result in a very different design process – from data collection extracted from the interaction to social analyses in real-life settings, accounting for the changes in the Human-AI relationship may shape a very different practice.

It is not enough to ask ourselves what we know about society and try to make sense of this knowledge in the practice of designing AI. It is also not enough to pursue different technological frameworks that are more fit for the vastness of the social world. This separation would only continue to limit the combination of effects that can only happen when designing AI not by holistically addressing technological and social concerns, but by addressing the merger between the two, i.e., by practising AI design as part of a large, complex and emergent socio-technological phenomenon.

To avoid this limiting effect, design thinking might prove a productive approach to bridging these disciplinary gaps, materialising into new, experimental modes of practice that do not fall into the trap of insufficient scrutiny (Cross, Hortensius and Wykowska, 2019). This could be achieved by pursuing the questions that embody both AI's technical conceptualisation as well as its ramifications on human social cognition when perceiving, communicating and cooperating with social agents.

In tackling this gap, this research aims to explore this intersection without specifically alluding to any particular social challenge as an object of study. Instead, its objective is to discover how these concepts are brought into existence through the Human-AI interaction and meaning-construction.

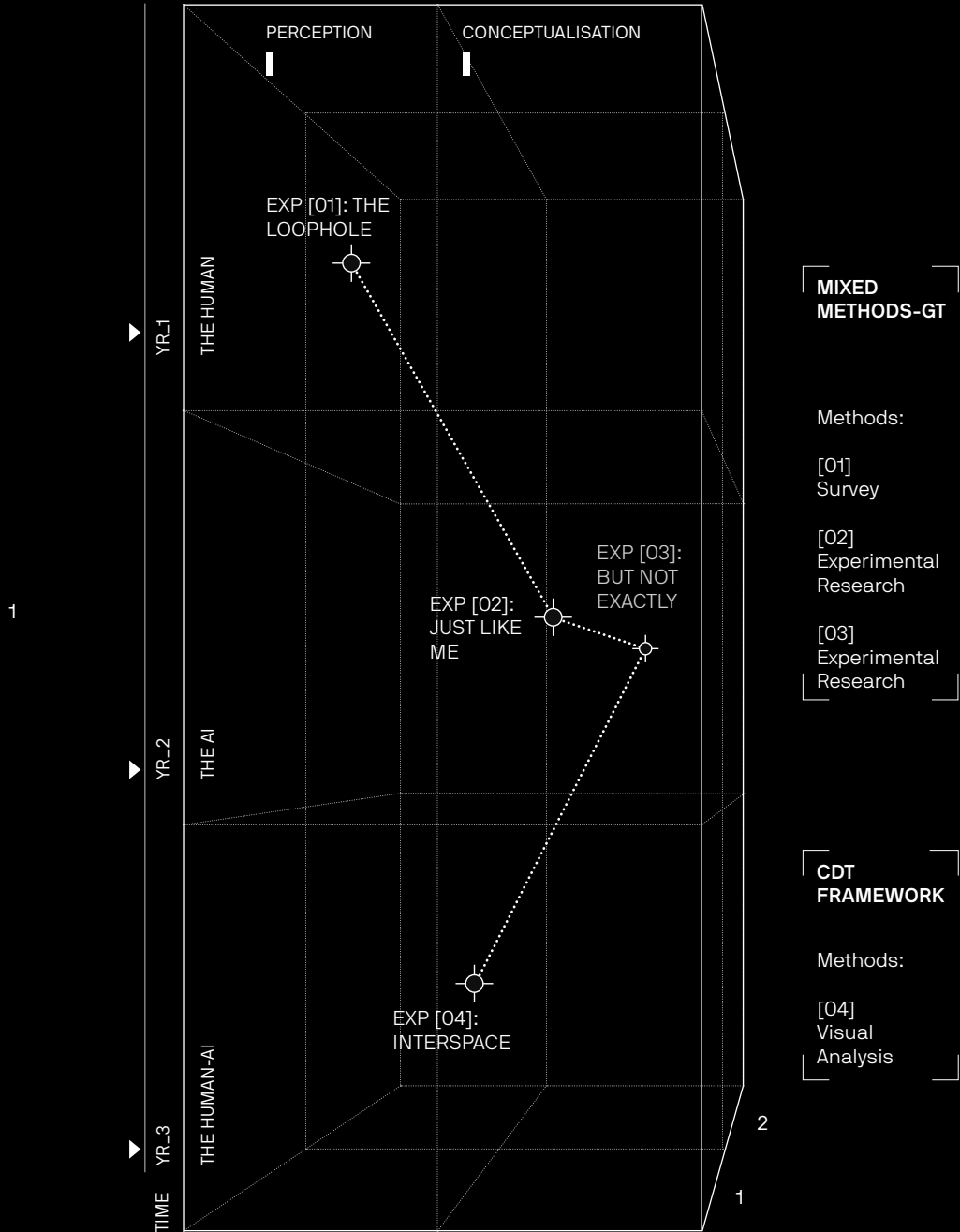
In doing so, it is trying to answer the following main research question – *How can design research navigate the gap between AI's social conceptualisations and human expectations?* – with three corresponding sub-questions that further deconstruct these two main variables with respect to the anthropomorphic phenomenon:

- + Is human behaviour driven by mental heuristics independent of a cognitive theory, or is it directed by more precise social schemas? (Chapter I)
- + Are anthropomorphic expectations considerably different from their human-reality counterparts? (Chapter II)
- + How are expectations continuously constituted through the interaction between the two in practice? (Chapter III)

This focus on an anthropomorphism, which is at the heart of this body of work, has been informed not only by the inevitability of a point of convergence between the social and technological factions of AI systems, but also by a preliminary experiment on the nature of self-representation within various technological systems. In this initial ‘sensitising’ study, further described in the Human chapter, it was noted that people made anthropomorphic inferences even in the absence of clear-cut humanlike affordances and that expectations of ‘smart’ technologies attested a penchant for anthropomorphic attributions. It was this initial insight that has informed not only the above research questions but also the motivation for choosing as an object of study the AI application that has been an integral part of the digital revolution: IoT and home automation; a technology with perhaps some of the broadest cultural, political and socio-economic repercussions.

STRUCTURE

In navigating this research question, I had to first delineate the two main variables into two distinctive investigative strands: conceptualisation and perception (Figure 1). In other words, in research, this entailed a distinction between conceptual versus mental models, or how we design an artefact or system versus how we actually perceive and interact with it. In practice, this implied having to make assumptions about the influence these variables have upon each other and deciding on their manipulation in order to answer the research question. To address this, based on existing research in this field, it seemed, at first, counterintuitive to pursue manipulating perception as a variable conditioned by the artefact’s design. This would only reinforce the common practice of regarding AI as a purely technological object, thus removing it from its social consequences until post-implementation.



I, therefore, started from the assumption that when dealing with the human interaction metaphor in AI systems perception would play a crucial role – one which is, to an extent, independent from the interaction. This hypothesis is not new but instead derived from an availability heuristic that has been extensively investigated in computer science research. From the dawn of personal computers, thinkers such as Turkle (2005) or Reeves and Naas (1996) explored how people resorted to mental shortcuts of treating computers as humans even with media that was never intended to display social capabilities in the first place. It is known, from these studies, that human heuristics may subliminally manipulate our perception of technology, but what could we expect to happen when intentionally designing technology to appear humanlike? Would observed behaviour be driven by mental heuristics independent of a cognitive theory, or would it be directed by more precise social schemas? Moreover, if the latter, what are some of the independent factors that influence anthropomorphism as a psychological phenomenon? This first analytical perspective focuses on the Human (Chapter I).

Next, once establishing perception's partial role in the dynamic between these two variables, I proceed to invert the analytical perspective by focusing on conceptualisations. In doing so, the social AI becomes understood along the lines of the assortment of humanlike affordances, with unknown implications on human perception. How would these affordances lend to the perceptive process? What type of mental models could we expect to find? Would they be entirely different from their human-reality counterparts? In establishing the degree of fidelity these mental models exhibit when compared to the reality of the social world, the discussion broadens to also encompass design's power to perpetuate stereotypes and deeply engrained social beliefs through technological interventions. This second analytical perspective focuses on the AI (Chapter II).

Lastly, I conclude the third chapter by suggesting that these analytical perspectives serve to demonstrate an epistemological limitation. The locus of the interaction could not be extracted from manipulations of these variables' effect on each other, but rather by looking at how it is continuously constituted through the interaction between the two, in practice. Therefore, this third analytical perspective focuses on the interaction as the unit of analysis emerging from two reciprocal (rather than dependent vs independent) variables. With this reconfigured frame of reference, it becomes more apparent that the practice of designing digital agents requires additional methodological approaches that could better serve the complexity of this nascent socio-technological world. The last analytical perspective focuses on the Human-AI (Chapter III).

METHODOLOGICAL UNDERPINNING

'How can design research navigate the gap between AI's social conceptualisations and human expectations?' – when considering this research question against the backdrop of the gaps in academic literature, the most pertinent strategy to address what seems to be a little known phenomenon lacking theoretical framework is Grounded Theory (GT) – a structured yet flexible research design aimed at producing or constructing explanatory theories of human attitudes and behaviours (Chun Tie, Birks and Francis, 2019).

However, Grounded Theory comes in many flavours that, despite arising from the same root and sharing a number of methodological strategies, are nonetheless neither consistent nor interchangeable (Kenny and Fourie, 2015). There are multiple points of divergence between genres, some of which appear to be poorly understood and in contradiction. The ones that require clarification moving forward are linked to (a) their opposing theoretical and philosophical perspective (b) contested research procedures and use of methods and (c) their contrasting use of literature.

Nevertheless, before commencing, it should be noted that the choice of methodology has not been entirely predetermined from the onset. Instead, it evolved in response to many factors, including ontological and epistemological perspectives as well as complex research needs that inevitably informed the unfolding of the practice. As a result, this section serves to establish the aspects that informed the genre of GT employed throughout this research rather than an engagement with the methodological framework and the research process (both of which will be further discussed in the ensuing chapters as they unfold in relation to the experimental practice).

Because of the high number of permutations resulting from tailoring the above-mentioned three variables in response to the complex research needs of this study, the shape of GT employed in this work could be seen as “hybrid design – or mixed-methods GT (MM-GT). The remainder of this section will concentrate on the three distinguishing areas, demarcating how these variables contributed to the “hybrid” research design. These perspectives and their relation to the choice of methodology are clarified below.

A. EPISTEMOLOGY / PHILOSOPHICAL PERSPECTIVE

GT consists of three prevailing traditions: classic or Traditional GT, associated with Glaser; evolved or Straussian GT, associated with Strauss and Corbin; and Constructivist GT, associated with Charmaz (Kenny and Fourie, 2015). While each flavour is an extension and development of the original GT developed by Glaser and Strauss, the main incongruity between these three approaches essentially hinges on their opposing philosophical perspectives. As such, defining the particular choice of methodology is more than a choice of research design. Rather, it is something that reaches far into the assumptions about the reality brought to this body of work. In this vein, I will first lay down the nature of these assumptions in order to justify the choice of research.

There is still ample debate over which paradigm Traditional GT best corresponds to, as the original textbook was virtually silent on questions pertaining to epistemology and ontology (Bryant, 2002; Urquhart, 2003). Still, a number of academics addressed this ambiguity directly. Charmaz, for one, reasoned that, despite Glaser's reservations, Traditional GT seems intrinsically correlated with positivism, as it implicitly assumes "an objective, external reality, a neutral observer who discovers data, reductionist inquiry of manageable research problems, and objectivist rendering of data" (Charmaz, 2000, p. 510). Even more, Strauss, one of the original co-founders of GT, acknowledged the positivist nuances embedded within the terminology of a theory that emerges from "out there, so corresponding to the ontological presupposition of an external, objective reality" (Corbin and Strauss, 1990, p. 279), waiting to be *discovered rather than constructed*. These assertions of "an external reality" that is nevertheless "independent from an observer and the methods used to produce it" (Charmaz, 2000, p. 513) are key positivist tenets that Glaser himself recognised. In response to Charmaz's classification of the objectivist epistemology embedded within classic GT, Glaser contended that his methodology serves to "make the generated theory as objective as humanly possible" (Glaser, 2002). However, while this assessment of positivism and post-positivism as being demarcated by a critical realist rather than realist ontology is still subject to debate, the assertion of objectivism remains.

The ontological presupposition of an external objective reality remains as we progress to the evolved GT or Strauss and Corbin's rendition of the classic theory – as an evolution from the classic or Traditional approach (Kenny and

Fourie, 2015). Despite Strauss and Corbin's more nuanced approach, incorporating participants' perspectives and the acknowledgement of these multiple perspectives, their meticulous coding conventions and methodological procedures serve the same original goal – that of striving for impartial data collection and analysis (Charmaz, 2000). These nuances stem from Strauss' ties to pragmatism and symbolic interactionism, a sociological perspective that addresses the meaning individuals place on objects, behaviours and phenomena based on social constructs and on what they hold to be true. Indeed, the founders of evolved GT located their approach within the theoretical perspective of symbolic interactionism, albeit still endorsing a realist ontology and a post-positivist epistemology (Aldiabat and Navenec, 2011). With this theoretical perspective, they argue that “although reality exists to be uncovered by inquiry, it is never perfectly apprehensible” (Guba and Lincoln, 1994; Ghezaljah and Emami, 2009, p. 17), a view that is consistent with a post-positivist epistemology.

Lastly, this thread of endorsing the principles of symbolic interactionism can also be found in Charmaz's Constructivist approach, albeit with some notable differences. It is precisely this different flavour of symbolic interactionism that informed the choice of methodology utilised in this research.

Symbolic interactionism, as a theoretical perspective strongly bracketing the choice of methodology in relation to this research question, has been defined by Herbert Blumer (one of the founding fathers of the approach) as being underlined by three simple principles (1986). First, he argues that “human beings act towards things on the basis of the meanings that the things have for them” (Blumer, 1986, p. 2). Secondly, he defines ‘meaning’ as not intrinsically present “out there” but as ascribed to actions, gestures or objects through social interaction (Blumer, 1986, p. 3). Lastly, and perhaps most importantly for the purpose of this investigation, Blumer posited that these meanings are always subject to fluctuations as they are redefined through the interpretative processes of social interaction.

Some of these very crucial aspects of fluctuation and interpreted meaning have not found their way into Straussian GT, which prioritised the resultant action rather than the meaning-making process. This led Charmaz to formulate a new approach that stood against Straussian's post-positivist, overly-prescribed regulations, which, she believed, impinged on the creative reflection and analysis (Charmaz, 2013). As a result, Charmaz proposed recovering symbolic interactionism's accent on language, interaction and meaning construction through a

more interpretative methodology deeply rooted in a constructivist epistemology. Constructivist GT thus reclaimed the tools of GT from their objectivist origins to forge a more flexible and open-ended methodology, this time unambiguously underlined by a constructivist epistemology (Bryant and Charmaz, 2019). Thus, we notice that language as a primary vehicle of socialisation and social interaction as the process of co-construction of meaning with socially-able agents, are approaches integral to the assumptions brought into this body of work. These considerations alone (if not for ample recognition that truth claims are epistemologically consistent with a constructivist stance and that they are by no means believed to be absolute, objective or truly generalisable, but rather, above anything, interpretable) are necessarily in line with a GT demarcated by a constructivist epistemology.

B. MIXED METHODS

Secondly, another key point of contention pertains to the debate on whether Grounded Theory is inherently wedded to certain methods and research approaches. Traditional GT and Straussian GT assert that GT is to be seen as a fully inductive methodology focused on theory-creation “through the systematic collection and analysis of data” (Corbin and Strauss, 1990). This consideration alone renders traditional GT as a fully exploratory/no-hypothesis approach, most frequently associated with qualitative research methods – despite the authors never formulating any such strict-qualitative requirement to begin with.

In other words, there is plenty of scope for both quantitative and qualitative methods – i.e., data collection – to be situated alongside a GT methodology (see Johnson, McGowan and Turner, 2010; Johnson and Walsh, 2019). Accordingly, mixed-methods GT has evolved as a popular methodological approach, which not only sees GT as a hybrid design that can be tailored to work with any method, but also challenges its very prescriptive, inductive origins. MM-GT often sees theory generation (induction) and theory-testing (deduction) as fully compatible, “allowing the mixing of (a) pure exploration/discovery and some a priori guiding structure, (b) qualitative and quantitative data, (c) idiographic/local- and nomothetic/general explanation, and (d) theory discovery/generation/ construction and theory confirmation/testing” (Shim, Johnson and Gasson, 2020).

This position is consistent with this study's requirement that it produces an explanatory theory for a social phenomenon following an approach that could not have been predetermined from the onset, permitted to evolve in response to the demands of the research requirements and thus allowing for both qualitative and quantitative data collection, as well as both theory testing and theory generation.

C. USE OF LITERATURE

Lastly, having established the flavour of methodology employed in this work, the discussion should revolve more around MM-GT's position on this matter. However, in order to define this position, we need to contrast it to the Traditional GT practice, a discussion that brings us back to their competing philosophical perspectives.

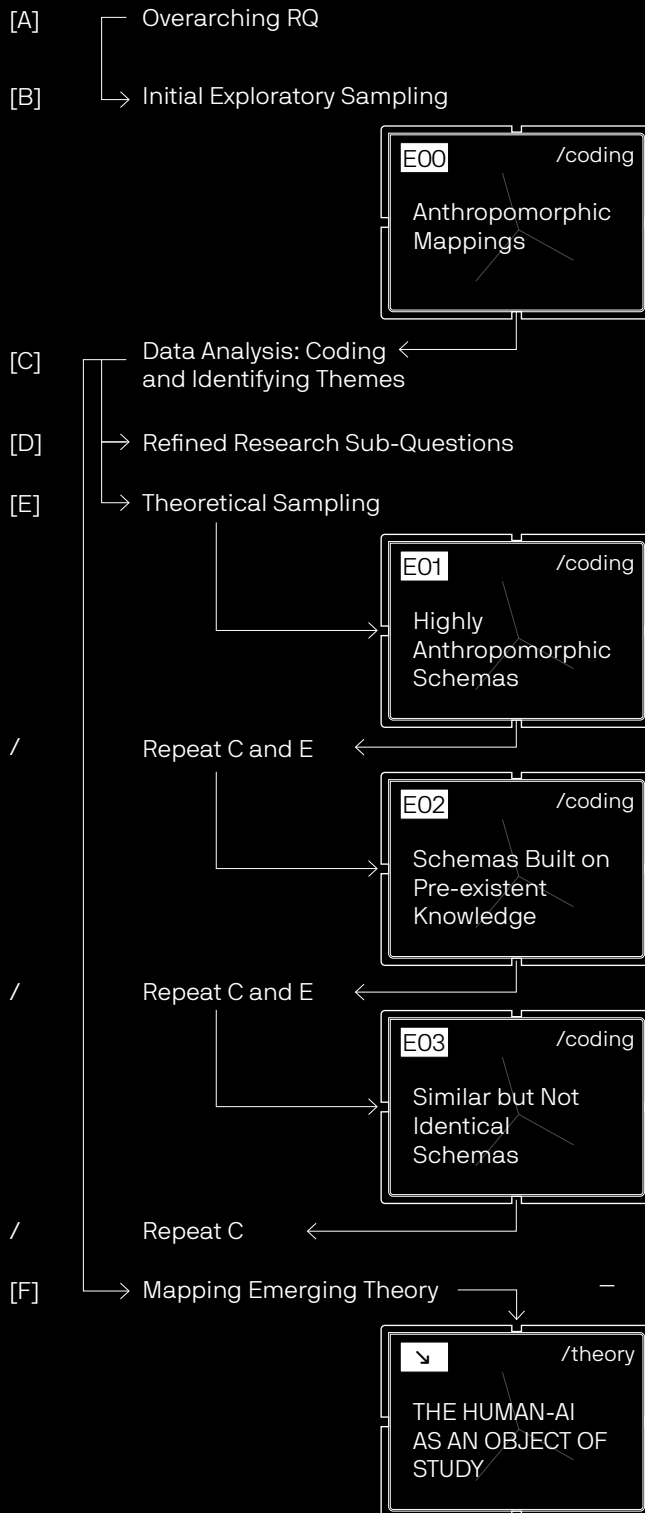
In the classic form of Grounded Theory, Glaser strongly asserts that any prior knowledge "violates the basic premise of GT", as it could contaminate, inhibit or stifle the analyst's ability to produce theory based on a dynamic new concept not previously mentioned in literature. The researcher is meant to suspend any prior knowledge and to refrain from the need to consult any academic literature in order to ensure an open mind, free of gratuitous influences. This position stems from the positivist's prerequisite of removing the researcher from the research for the sake of objectivity (Kelle, 2005).

However, there is a perceptible difference between an *empty head* and an *open mind* – an argument with which both subsequent flavours of GT challenge this traditional stance (Corbin and Strauss, 1990; Kelle, 2005). Consistent with a constructivist paradigm that accepts a researcher's influence as inevitable, the subsequent forms of GT both accept prior exposure to the subject as well as to literature, in all phases of research. Strauss and Corbin even argued that consulting the academic literature yields many benefits such as: revealing gaps in academic knowledge, inspiring novel questions, guiding theoretical sampling, use as additional validation and providing further insights into the existing theories and bodies of knowledge. This position is consistent with the assumption in constructivist philosophy that no research occurs in a vacuum, and it is always inherently influenced and informed by the contexts in which it operates (Charmaz, 2013).

In echoing these outlined methodological strategies, the ensuing three chapters have not been predetermined from the onset of the study, as perhaps might have been expected with certain types of GT research. The MM-GT design emerged as a way of tackling the overarching research question by breaking it down into three broad phases (Figure 2). However, the decision to jump from the Human to the AI and then to the interaction between the two could not have been prearranged. Instead, it emerged from an analysis that revealed further gaps relevant to the process, which necessitated additional data collection and hence the need to change the direction of study. This is to be expected from a GT approach (Kenny and Fourie, 2015).

As a result of this progressive sampling guided by unfolding identifications, predetermining the type of methods or literature from the onset of the study would have been not only impossible but counterproductive. Further, this research will not be presenting the reader with either a comprehensive literature review chapter, nor a chapter on methodology beyond what has already been discussed above. Instead, every research theme (the Human, the AI and the Human-AI) will have its own consultation with literature, as well as its own discussion on the specific research design choices and methods employed.

Echoing the considerations mentioned above, I began this investigation by delving into the culture of one of AI's most distinguishable applications, which has been not only an integral part of the digital revolution but also an epitome of AI as a merger between its artefact and phenomenon implications: Internet of Things in the context of home automation.



01
THE HUMAN

PROLOGUE

As perhaps is to be expected, the ensuing investigation did not commence from the very specific research question mentioned in the introductory chapter but arrived there following an initial stage of data collection and theme identification. Significantly, this research began in 2015 as a fully funded proposal between the Royal College of Art and Microsoft Research Cambridge aimed at examining a cluster of technologies more commonly known as the Internet of Things (IoT). The motivation for this practice-based research concerned the roles and responsibilities designers have in shaping this emerging technological space, providing insights pertaining to the uncertainties engulfing much of their socio-cultural impact, and the need for more appropriate design methods and practices.

The Internet of Things has emerged as a significant technology – seen here, in line with Madakam, Ramaswamy and Tripath (2015), as an ‘AI subtechnology’ – that shapes digital culture by connecting physical devices / embedded objects / ‘things’ to existing internet architectures. This initial goal of connecting various sensors and actuators with the purpose of collecting and analysing data to generate situational awareness is reminiscent not only of the 1930s and 1950s hope for the “homes of tomorrow” (Spigel, 2001) but also of our 1950s dream of enabling machines to make sense of their environment.

The attraction of the Internet of Things continues to hypnotise. No one doubts anymore the fact that society’s ever-increasing reliance on digital technology has inevitably taken grip of the way we process and exert many of our societal practices, including the practice of design. This informational technology has provided a means of truly exploring a realm that can exceed previous notions of reality and convention, a change that inevitably leads us to a new type of product in which the informational and the material, the virtual and the actual, the possible and the real, become more and more interwoven. As a result, we see a distinctive design culture being born, with its own values, conventions and meanings, leading to ideologies that sometimes undermine and sometimes marginalise the thinking systems of the real world. With ideological discourses serving the prevailing metaphors of this emerging culture, the initial research question that informed the direction of this study was related to its impact on the design practice and design’s integration of the Human in relation to this emerging ‘product’.

In this vein, this investigation commenced with a preliminary study on IoT – an AI subtechnology that has managed to find its way into the most intimate of spheres: our homes. Inspired by Tom Igoe and Dan O’Sullivan (2004) illustration of “how the computer sees us” (represented as a simplified human that has one gigantic finger because we communicate through sequential tapping, one eye instead of two, because we stare at 2D screens, and two small ears to listen to stereo sounds), eleven participants from the Royal College of Art had been invited to draw representation of how they presumed a series of ten different IoT devices represented them. The ensuing drawings were then analysed for content, style and structure prior to being aggregated into single cognitive representations of the elements measured as most salient.

Contrary to the initial expectation of finding inadequate self-representations withing a technical system, the resulting diagrams were – regardless of the complexity of the computer – noticeably anthropomorphic; represented as more or less abstracted versions of the self. The presence of humanlike mental models even in the absence of clear-cut humanlike traits illuminated the relevance of anthropomorphism to this study’s inquiry and design research space. As a result, this investigation commenced with a brief investigation into the culture of automation and IoT with respect to the anthropomorphic phenomenon.

IOT <> AI

In 1991’s *Scientific American*, Mark Weiser, then Chief Technologist at Xerox PARC, started his seminal paper with a statement that was to influence a generation’s perception of the role technology plays in our lives: “The most profound technologies are those that disappear. They weave themselves into the fabric of everyday life until they are indistinguishable from it” (Weiser, 1991). His quote, which spoke to technology’s ultimate purpose – that of being an invisible extension of a more empowered self – ended up becoming a tenet of technology design. As such, in achieving the ambiently intelligent world that responded to people’s presence, IoT soon met AI, and in blending AI into the fabric of our everyday life, figuration emerged as a potent design tool. In this section, I will first look at the ways in which this web weaves itself into our social structures through the use of AI, followed by a discussion of how, in achieving the goal of “indistinguishability”, AI coalesced with another powerful design technique: anthropomorphism.

There is not one single accepted definition of what the Internet of Things represents; instead, definitions seem to converge towards one common denominator: the idea that the first form of the Internet was populated with data produced by people, whereas Internet 2.0 became about data generated by things. As such, Madakam defines the Internet of Things as “an open and comprehensive network of intelligent objects that have the capacity to auto-organize, share information, data and resources, reacting and acting in [the] face of situations and changes in the environment” (Madakam, Ramaswamy and Tripathi, 2015).

In *Dangerization and the End of Deviance*, Lianos and Douglas (2000) argue that the development of Internet 2.0 has been primarily motivated by corporate desires to benefit from the predictability “afforded by the ability to follow objects and resources through the commodity chain in which they are embedded”. IoT has thus enabled companies to track and analyse objects to increase efficiency, reduce errors and speed up a process fully embedded in an organised system of things. The word ‘things’ here has always been used vaguely, coming to stand for anything that is distinguishable in the real world. That is, not only home appliances and gadgets, but things, both living and non-living, that we do not normally think of as electronic at all. From materials and equipment, merchandise, food, clothing and furniture, landmarks and art, to animals, plants and fruits, at this point, the word ‘things’ has come to represent anything with a presence in our real, material world that could be linked to this vision of a global infrastructure of physical objects. Essentially, in this global infrastructure, sensors and actuators embedded in real, physical objects are connected through wired and wireless systems, which often use the same IP address to connect to the Internet. In doing so, these networks generate immense volumes of data that, to be interpreted correctly and respond swiftly in accordance with the complexity of the real world, require analysis.

Enter AI. With the increasing number of sensors, high-speed networks and workstations, Artificial Intelligence remains the best solution for both data mining and management of the IoT network. Even more, AI has become the perfect environment for achieving that subconscious design goal of providing a human-computer interface that vanishes in the background of our daily lives. In supporting people to naturally carry out their activities, AI in the context of IoT has come to stand for characteristics such as: context awareness (embedded in devices that recognise the Human and their situational context); personalisation

(tailoring responses to the Human needs); adaptiveness (changing and learning in response to the Human) and anticipation (anticipate desires). Since all these system characteristics allude to nature-inspired capabilities, the paradigm of the human trope soon found its way into the design of a more 'natural' human-IoT interface.

AI <> ANTHROPOMORPHISM

On the 6th of November, 2014, a year prior to commencing this research, Amazon stunned the whole tech community with their latest “futuristic gadget” – the Amazon Echo. The company chose to forego the customary fanfare of tech releases and instead announced the Echo with a press release comprised of a series of short promotional videos showcasing a minimal and shiny black cylindrical object with animated blue-green lights. While aesthetically it closely resembled an already popular product, a wireless speaker, its interface soon became indicative of the perfect convergence between the Internet of Things, AI, and automation. With an in-built digital assistant named Alexa which could react to verbal commands, it rapidly became apparent that this sleek device was able to provide much more than simply a means to play music. Amazon’s string of short promotional videos showcased Alexa providing information about traffic and weather, synchronising with other applications to schedule reminders and appointments, creating shopping lists and, more importantly, connecting to other ‘things’ such as thermostats, light bulbs and switches to effectively function as a voice-activated “smart home hub” (Amazon.com, 2020).

As with any popular new technology, public reaction to Echo was polarised. On the one hand, it was being praised as exactly “what smart homes should feel like” (Wong, 2015), while on the other, the uneasiness of its surveillance capabilities led to it being described as “a trojan horse to penetrate our remaining private moments” (Wasserman, 2014). Fully equipped with a series of microphones designed to always listen, concerns over how its interactions were to be recorded and stored in a cloud, effectively representing a passively constructed digital self, unless manually deleted, soon took over.

Nevertheless, as one of the first products to integrate AI with voice control and home automation, Echo received virtually instant commercial success, coming to dominate, over the next few years, both the market of IoT devices and the

public consciousness as the exemplar of smart home technology (Phan, 2019). Everything about it had gone viral, from its initial commercials to spin-off parodies, to hundreds of online forums and communities dedicated to discussions on its features, to thousands of customer and tech reviews in press and publications such as CNET, PC Mag and Tech Radar. Following its initial US release in 2015, the Echo soon came to dominate both the UK and German markets by the end of 2016 and by 2017, it became one of the most popular devices of its kind, expanding into 34 countries (Consumer Intelligence Research Partners, 2017). By January 2019, an estimated 100 million Alexa-enabled devices found their way into homes all-over the world, a figure that included about 63% of the UK's home automation market (Skeldon, 2018).

However, Alexa's instant commercial success has been more than just the result of its home automation skills. With its release video depicting a fictional American suburban family, in a strikingly heteronormative scene of domestic life, unpacking the "newest member" – the scene was designed to provide a comforting reassurance of a technology that would only solidify traditional domestic ideals (Spigel, 2001). Echo's success has been, in part, the result of this figuration as an "idealised vision of domestic service" (Phan, 2019). The term personification or figuration means here the act of generating a representation of a person – a representation that does not automatically correlate to its 'real' settings, but rather to the imagined model imprinted socially.¹ Of course, this is not to say that it does not have real implications, which, as it turned out from later experiments, it does.

And so, we come to anthropomorphism – an effective design strategy, which in Phan's account (2019), romanticises the relations of "servitude", making it much easier to accept a misrepresented power dynamic between the user and the device – its figuration as a household aid, in fact, cleverly obscured issues such as "hierarchical surveillance and digital labour", leading to more easily tolerate security and privacy concerns.

As such, in my contention, the use of the human metaphor emerged as arguably one of design's most potent tools for the commodification of our most intimate selves. With anthropomorphism eliciting something, I believe, to be more than a

1 It should be noted the distinction between *anthropomorphic features* (figuration) and *anthropomorphic perception* (anthropomorphism) – terminology which will be further clarified in the following pages.

mere social reaction, the remainder of this chapter will focus on unpacking the consequences of personified design on human cognition and discuss the repercussions of invoking the anthropomorphic perception when designing a cluster of technologies more commonly known as Smart Assistants, Intelligent Personal Assistants or Personal Virtual Assistants.²

ANTHROPOMORPHISM <> DESIGN

The human metaphor has traditionally been seen as one of the most obvious design strategies for superimposing aspects of the human observer onto objects and interaction. Lamps, radios, food products, branding, vehicles, software and computers: anthropomorphism ranges from the conspicuous to the elusive, from aggressive to subversive. Anthropomorphism – from ancient Greek *ánthrōpos*, (“man, human being”) and *morphe*, (“form, shape”) – is defined as the act of attributing humanlike qualities, such as intentionality and mental states, to nonhuman agents (*‘anthropomorphism, n.’*, 1885). In a world with a staggering diversity of living, spiritual and technological agents, the uniformity of our mental representations of them is suspiciously familiar. Animals imbued with motivations, intentions and plans; gods with familiar appearances, instilled with personality weaknesses and strengths; rogue computers that plot against us, fickle printers, capricious weather and intentionally erratic fate – they all epitomise a recognisable human tendency of infusing real or imagined nonhuman agents with humanlike motivations, intentions or emotions (Guthrie, 1995). As a result of this readily available interpretative model, the anthropomorphic paradigm has managed to penetrate numerous social practices ranging from marketing strategies to theories of religion and, of course, design.

Design culture has incorporated the anthropomorphic formula from its origins. Tracing back to behavioural modernity, the plethora of humanlike pottery or human-shaped tools are early attestations of human’s affinity for designing to elicit anthropomorphic reactions. Correspondingly, as the practice of design

² Within this context, the process of data sampling and analysis began as a means of probing the evolution of our anthropomorphic perception of these emerging technologies and whether the extent of perception can be merely predicted by the theories of anthropomorphic cognition in practice (i.e., detached from the specifics of their design affordances, which will be discussed in more detail in the following chapter).

expanded and matured, anthropomorphic elements have remained a persistent leitmotif, pushing its application well beyond acts of embellishment. Akin to its initial literary function – that of a storytelling device – the practice of projecting values or intelligence to rationalise a system’s appearance or behaviour has acquired a privileged position in the designers’ repertoire of tools. To illustrate some of the avenues in which this is has been achieved, DiSalvo and Gemperle (2003) identified four different design incentives behind this strategy:

1. Archetypal products that have long utilised the anthropomorphic form strive to maintain conventions and avoid confusion over identity or function. This aspiration for uniformity and consistency is most distinctly exemplified in the design of vehicles or domestic robots, which have conventionally been designed to incorporate humanlike features.

2. Anthropomorphism has been utilised as a powerful resource for eliciting the projection of personal, social or cultural values in association with the features or activities provided by a product. In this instance, we notice how a vehicle’s choice of joyful or angry headlights taps into that value’s appeal to very specific demographics (Figure 3 & Figure 4).

3. The anthropomorphic form has seen considerable usage in mirroring a product’s attributes or functions, mainly with the intent of structuring our relation and interaction with it. This is more recognisable in perfume or soap bottles, which customarily take the shape of a female torso as an indication of not only the associated sensuality, but also to suggest their usage is mainly intended for women (Figure 5 & Figure 6).³

³ It should be noted that anthropomorphism can be seen as a subset of animism – which is the cognitive inference that inanimate objects are alive, whereas anthropomorphism is the cognitive inference that inanimate objects are human (Karpinska-Krakowiak and Eisend, 2021). The terms share certain similarities in terms of perceptual mechanism, yet unpacking the differences goes beyond the scope of this research. I will similarly not be discussing vaguely related concepts such as zoomorphism – the converse of anthropomorphism, i.e., the act of attributing animal-like mental states to humans; or skeuomorphism which is the general practice of mimicking real-work counterparts in design. While they do share some similarities in practice, the relation between them does not concern the nature of this investigation.

3 Canepa Motorsport (2020),
1959 Austin-Healey Sprite
© Rights Canepa Motorsport,
all rights reserved.

4 Canepa Motorsport
(2022), 20 Ford GT 6634
© Rights Canepa Motorsport,
all rights reserved.

3



4



5 Pieters, Geert (2020) Jean Paul Gaultier Classique Bottle
Available from: unsplash.com

6 Spaa, Joppe (2020) Azzaro Chrome Eau de Parfum for Men
Available from: unsplash.com

01



5



6

41

4. What is perhaps the fundamental consideration in debates over anthropomorphism's appropriateness in design, is its utilisation as a means of explaining the unknown by providing intuitive and readily accessible mental models of interaction. Dennett (1998) best describes this utilitarian function as the "intentional stance", whereby unintentional agents, such as IoT devices, computers or robots, are attributed humanlike intentions to increase the ease with which people can reason about how they function (and thereby interact with them more efficiently). Proponents of the utilitarian function argue that anthropomorphism is an essential tool in increasing the comprehension and predictability of what would otherwise be indistinct machinery, and a vital aspect in achieving and maintaining sustainable relationships with our technological products in general.

I have, finally, added two more arguments to the original list, which, in my view, are more closely related to this topic of investigation:

5. Indistinguishability – as a result of the aforementioned technological aspiration towards full integration into our daily lives. Motivated by various experiments that exposed people's instinctive propensity to extend specific social considerations to computers based on minimal cues (see Reeves & Nass, 1996), computer science fields have also argued for further incentives, which have subsequently been quickly adopted in design practice.⁴ The very fact that users unconsciously perceived computers as social agents denoted an obvious opportunity for the human-interaction metaphor to find its place in interaction design (Fineman, 2004). Furthermore, Kremmentsov and Todes (2010) even commented that given "the long history of anthropomorphic metaphors... may testify to their inevitability". As such, the general consensus, especially in the field of Human Robot Interaction (HRI) (Duffy, 2003; Duffy and Zawieska, 2012) and Human Computer Interaction (HCI), is not how to avoid anthropomorphism but rather how to better embrace it (Duffy, 2003).

6. Social integration – amongst some of the crucial implications of anthropomorphism, especially in HRI and AI, lies our tendency to render humanlike agents worthy of moral care and consideration (Waytz, Epley and Cacioppo,

4 The Media Equation, developed by Clifford Nass and Byron Reeves at Stanford University, states that people's responses to media are fundamentally social in nature, reacting to different media as they would to other people.

2010).⁵ The repercussions of this can be seen in instances whereby a strikingly human resemblance and demeanour help machines gain acceptance in the public sphere or even acquire legal personhood. Thus, it does not come as a surprise that, in *Dimensions of Mind Perception*, Gray, Gray and Wegner (2007) reason that understanding the process of anthropomorphism is central not only to multibillion-dollar industries such as robotics but also to debates on agency and morality, ranging from environmental consciousness to abortion rights.

The use of anthropomorphism is, however, not straightforward. Part of the ensuing need for experimentation lies in the fact that these justifications have been the subject of ongoing debates, ranging from those in favour of anthropomorphism as an effective method of facilitating social interaction (Duffy, 2003; Duffy and Zawieska, 2012) to those that argue against false expectations and dangerous relationships with technological products (Shneiderman, 1989; Don *et al.*, 1992). At this latter end of the spectrum, researchers such as Shneiderman view the role of anthropomorphism in HCI as a design compromise that leads to issues of unpredictability and vagueness – issues that go directly against the importance of comprehensible, clear and predictable interfaces that support direct manipulation. In his influential paper on anthropomorphism and robotics, Duffy responds to this view by stating that “Shneiderman’s comment touches on a problem which is not fundamentally a fault of anthropomorphic features, but a fault of the HCI designers in not trying to understand people’s tendency to anthropomorphise, and thus they indiscriminately apply certain anthropomorphic qualities to their design which only lead to user over-expectation and disappointments when the system fails to perform to these expectations” (Duffy, 2003).

It should be noted that in this chapter, on the Human, anthropomorphism has been used to strictly denote the perceptive process, that is, the human rationalisation of a system through superimposing aspects of the self. This definition

5 Similarly, anthropomorphic traits have been used extensively to appeal to our core human nature (DiSalvo and Gemperle, 2003): on one instantly recognisable illustration, the recent spurge of ‘adorable’ or ‘baby-like’ design, such as Alessi house-ware range, is a confirmation of our attraction towards products that spark social or nurturing instincts (Marcus, 2002).

functions as a separator between *anthropomorphic features* (referred to earlier as figuration or personification) and *anthropomorphic perception* – terminology that, to add to the confusion, is far from consistent in research literature and computer science texts alike. This required further clarification.

Anthropomorphism, within social sciences, denotes a process of inductive inference about nonhuman agents, an inference that would be no different from similar inductive processes. In my contention, however, this term stands in contrast to HCI's or design's common utilisation of the term, which sometimes alludes to the deliberate act of designing anthropomorphic features, other times to the perceptual aspects of the phenomenon, and occasionally to both. This interchangeability perhaps furthers a lack of consensus on the topic, as it might subtly entail that the mechanism through which something is personified by design, directly juxtaposes with the user's anthropomorphic perception, which of course, is not necessarily the case. To clarify the use of these two very different notions, here, the use of *anthropomorphic features* will be referred to as 'personification', while *anthropomorphism* will continue to denote the process of induction.⁶ This, in design, corresponds directly to the notions of conceptual model (personification) versus mental model (anthropomorphism).

In design, engineering and HCI alike, the conceptual model usually denotes a set of tools devised to help users comprehend what a system does. Its primary function – that of representing a system as an abstraction of items in the real world – is intended to facilitate useful interactions and comprehensions of the system (Gentner and Stevens, 1983). However, while conceptual models are conceived by designers or engineers as deliberate representations of a system, people nonetheless form naturally-evolving mental models that need not be technically accurate. Thus, mental models or schemas are what people *actually* believe or think a system does, which all too often dramatically differs from designers' deliberate conceptualisations (Baecker and Buxton, 1987). In the case

⁶ As an often-unclear distinction, anthropomorphism and personification have been primarily treated differently in classical rhetoric. Personification or prosopopoeia is identified as the act of representing, usually only figuratively, abstract concepts or objects with or through human traits. Anthropomorphism's use in figurative language is often understood as personification's more literal equivalent, applied as an actual humanisation of nonhuman agents. This distinction virtually disappears outside of classical rhetoric, where anthropomorphism is used to denote both the act of assigning human qualities to a nonhuman agent as well as the perceptive nature of the phenomenon.

of anthropomorphism, designers employ humanlike conceptual models of a system, but this does not necessarily lead to predictable mental models.

The relationship between the two, however, is not that evident, and while the use of personification through anthropomorphic features such as voice, facial abstractions or movements has been used to prime models of values of interaction, the general belief is that anthropomorphism or sociality are fundamentally observer-dependent (Duffy, 2003).

Consequently, to understand how design becomes complicit in the issues pertaining to our human-humanlike AI interactions, we must first understand the mechanisms underlying anthropomorphism – a mechanism that, in line with Duffy's aforementioned view, should provide the key to designing suitable sociable agents.

However, despite the pressing need to better understand people's tendency to anthropomorphise, very few experiments have rigorously examined the mechanism behind this perception. As a result, it could be argued that a range of issues associated with our Human-AI interactions are born out of design's unfamiliarity with anthropomorphic perception. According to sociologists Nicholas Epley, Adam Waytz, and John T. Cacioppo (2007), very little research on anthropomorphism tries to provide a cognitive justification for when and why people are likely to anthropomorphise nonhuman agents. Instead, the focus remains on examining the precision and functionality of these anthropomorphic portrayals. "Most existing research therefore addresses the actual mental states and abilities of nonhuman agents by detailing the extent to which these agents really are humanlike and the extent to which anthropomorphism represents accurate insight versus fanciful illusion". In their view, the existing literature on anthropomorphism is inclined to "overlook the very psychological phenomenon in its midst", by predominantly studying the accuracy of these beliefs rather than by providing a psychological clarification for the beliefs themselves.

As with any multifaceted cognitive phenomenon, anthropomorphism is influenced by a range of dispositional and situational factors that make this remarkably common phenomenon nevertheless not invariant (Waytz, Epley and Cacioppo, 2010). Ranging from 'as if' metaphorical reasoning to strong forms of anthropomorphic inferences – whereby agents are overtly considered to hold human characteristics and behaviours (such as in the case of religious agents) – some

nonhuman agents are unmistakably anthropomorphised more than others. This spectrum is reflected in the diverse use of distinctive mental models extending from nonspecific humanlike traits all the way to an all-encompassing attribution of person and stereotyping. These models, though distinct, are all equally typified by the presence of humanlike traits. As such, we should also naturally expect to find users' perceptions of AI agents to similarly vary along a continuum, ranging from some anthropomorphic beliefs held very strongly, to some held more weakly, analogous to how social schemas and beliefs (such as stereotypes) do not all exhibit the same degrees of strength.

Then, is the anthropomorphic spectrum “observer-dependent”? And, if so, what fundamental aspects of the Human must we understand before deploying personified conceptualisation? These questions emphasise the need to study the Human in its interaction with the social AI.

THE HUMAN <> ANTHROPOMORPHISM

Compelled by its substantial implications on people's choices and beliefs, this variability in anthropomorphism has encouraged a considerable number of researchers to explore theories that could predict the behavioural consequences that ensue. One such theory, developed by psychologists Epley, Waytz and Cacioppo (2007), seems to have taken the lead in the world of HCI and HRI. The literature on the topic reveals that The Three-Factor Theory of Anthropomorphism is commonly used in tandem with Haslam's theory on dehumanisation (Haslam, 2006) as a comprehensive tool of quantifying the extent of the anthropomorphic spectrum when designing personified technology (see Złotowski, Proudfoot, Yogeewaran, & Bartneck, 2015; Złotowski, Strasser, & Bartneck, 2014).⁷

⁷ Kelman (1973) first explored the concept of dehumanisation in the context of mass violence, focusing on situations that weakened usual moral restraints. According to him, dehumanisation involves the act of denying someone a personal identity – “as an individual, independent and distinguishable from others, capable of making choices” (p. 301). Haslam (2006), however, approached this topic by looking at the aspects of humanness essential to evoking compassion and moral emotions, and mapped them on two distinct scales. This theory of dehumanisation, therefore, yields some essential insights into the opposite process, anthropomorphism – which is the act of humanising agents.

The Three-Factor Theory of Anthropomorphism (Epley, Waytz and Cacioppo, 2007), developed in an attempt to “synthesise a widely dispersed literature that predicts variability in anthropomorphism across dispositional and situational factors”, proposes a model that maps anthropomorphic thought onto three psychological determinants – “the accessibility and applicability of anthropocentric knowledge (elicit agent knowledge), the motivation to explain and understand the behaviour of other agents (effectance motivation) and the desire for social contact and affiliation (sociality motivation)”.

In summary, this theory suggests that the presence and variability of the anthropomorphic belief is dictated by three distinct determinants – comprising of one dispositional factor (elicit agent knowledge) working in tandem with two additional motivational determinants, ‘effectance’ and ‘sociality’ – each element making unique predictions about the strength and manifestation of the anthropomorphic perception.

A. ELICIT AGENT KNOWLEDGE

The first determinant, ‘elicit agent knowledge’ – referring to our egocentric tendency to reason about other humans or entities by utilising our own mental states and characteristics as a guide – can be justified by merely accounting for the impossibility of having access to any phenomenological experience other than being human. As we cannot know what being an animal, a deity, or a machine feels like, we are eventually bound to use our own mental states as a basis for induction (Higgins, 1996). “This invariant feature of sensory apparatus means that knowledge about human experience will be directly experienced and thus be acquired more easily, more completely, and more quickly than any knowledge (however indirect) about what it is like to be a nonhuman agent” (Epley, Waytz and Cacioppo, 2007). Thus, as our sensory experience is manifestly embodied, every agent’s subjective experience is bound to be forever out of reach. This is equally reflected in our inability to ever know what it is like to be any other person, leading to the trivial use of our own mental states as a starting point for induction when making sense of human behaviour in general (Nickerson, 1999; Epley *et al.*, 2004; Bahnik, Englich and Strack, 2017).⁸

⁸ Resulting in a universal bias of rarely ever being able to relate actions or beliefs precisely as another person would conceive them.

As an inductive process, Epley, Waytz and Cacioppo (2007; 2010) propose that anthropomorphism entails the mechanisms of any other process of induction – that of commencing with highly accessible knowledge structures (mental models) as an anchor, only to be subsequently corrected and redressed to acknowledge the framework of the nonhuman target. This hypothesis suggests that anthropomorphism – as an act of invoking a human mental model as a basis of induction – would ultimately be revised with the acquisition of knowledge structures that yield improved extrapolations about the agent’s behaviour. We are, however, required to consider an important caveat pertaining to the subject of this investigation: since knowledge about the self or other people, in general, is very likely to be “highly accessible and subsequently applied with little correction when reasoning about nonhuman agents” – anthropomorphic schemas might be challenging to correct, especially in contexts where attaining further knowledge is likely to be inaccessible or obfuscated by the nature of the technology, such as in the case of Blackbox⁹ technology.

Either way, the prediction that as our mental models expand to encompass knowledge structures of the target nonhuman agent, anthropomorphism is less likely to manifest to the same extent (as knowledge structures about humans would be less likely to form the basis for induction) should still be experimentally demonstrable. We should therefore expect to discover a moderation in anthropomorphism with time, as prolonged interaction with nonhuman agents would, theoretically, provide ample opportunity for familiarisation and correction. This initial hypothesis – which could yield thought-provoking insights for AI design – required further assessment within the context of home AI assistants.

B. EFFECTANCE MOTIVATION

The second factor, ‘effectance motivation’ – derived from White’s theory on the need to “interact effectively with [one’s] environment” – refers to the human desire of increasing the predictability and apparent controllability of the environment and the agents that inhabit it (White, 1959). Anthropomorphism, in this case,

⁹ “Device, process, or system, whose inputs and outputs (and the relationships between them) are known, but whose internal structure or working is (1) not well, or at all, understood, (2) not necessary to be understood for the job or purpose at hand, or (3) not supposed to be known because of its confidential nature.” (Bunge, 1963)

provides an “intuitive and readily accessible method for reducing uncertainty in contexts in which alternative nonanthropomorphic models of agency do not exist (such as those provided by science or culture)” (Waytz *et al.*, 2010).

This motivational factor, therefore, stipulates that in the case of AI agents, anthropomorphism is motivated by a need to increase the confidence and predictability of what would otherwise be unpredictable machinery, facilitating the ease with which we make sense of their actions. In this vein, anthropomorphism can be seen as something of a panacea offering closure and control, as a means of organising the present to establish predictability in the future. It should be noted, that this incentive closely aligns with Guthrie’s (1995) account of anthropomorphism as an evolutionary ‘safe bet’.

Intriguingly, this factor drives much of the rhetoric on the benefits of introducing humanlike traits in HCI, interaction design and robotics. Founded on our psychological need to counteract uncertainty and ambiguity in order to interact more effectively, proponents of this utilitarian function argue that people’s motivational determinant to anthropomorphise (as a means of providing a source of testable hypotheses) would ultimately provide readily available models of interaction. So, by converting this psychological determinant from motivation to justification, design practices have been broadly incentivised to utilise what Dennett (1998) best describes as “the intentional stance”.¹⁰

By aggregating these arguments into verifiable predictions, we should find that the anxiety associated with the unpredictability or ambiguity of an agent ought to ultimately incentivise anthropomorphic representations, or, in other words, when effectance motivation is high (such as when agents are novel, unpredictable, little understood or they violate one’s expectations), we would expect a stronger presence of the anthropomorphic schemas.

Moreover, Epley *et al.* (2007) also estimate that “agents one is expected to interact with in the future should be anthropomorphised more extensively than agents one never expects to see again”, as we would be, theoretically, more incentivised to interact efficiently. This, too, requires further scrutiny in the context of home AI assistants, whereby agents equally defy expectations or act unpredictably and noticeably form the basis of long-term interactions.

10 Whereby nonhuman agents are attributed humanlike characteristics and intentions, in order to expedite users’ ability to reason about them and thereby interact with them more effectively.

One closing aspect that must be noted with regards to users' effectance motivation towards social AI agents concerns the increasing unpredictability of acts of violating users' expectations, which is predicted to increase users' effectance motivation. This could ultimately imply that anthropomorphic representations are self-sustained by the mismatch between users' mental models and the AI's observed behaviour – especially augmented by the intrinsic inability to fully appreciate the underlying algorithms. This 'anthropomorphic loophole' is, presumably, rather problematic to test; it is, however, an aspect that must be considered when analysing and interpreting experimental data.

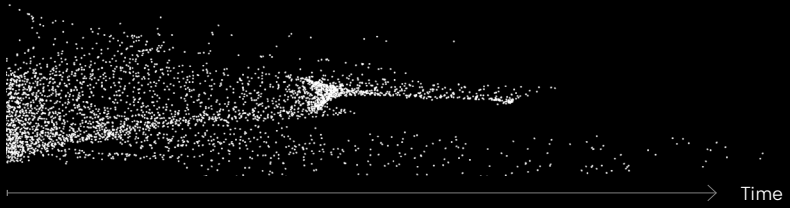
C. SOCIALITY MOTIVATION

Thirdly and lastly, 'sociability motivation' – referring to our desire to establish social connections with others, particularly when deprived of a sense of social inclusion – predicts that our sense of belonging will have an impact on our tendency to actively search for sources of social connections in our environment, and thus curve our predilection for anthropomorphism. This expected desire for sociability has also significantly contributed to much of the rhetoric on the benefits of personifications through design (Złotowski *et al.*, 2015). The necessity to establish and maintain a sense of social connection with others, and the perceptible ease with which nonhuman agents could satisfy this need, has notably facilitated the broad introduction of social AIs into the domestic sphere. As designers slowly uncovered a new avenue for alleviating social pain, more and more nonhuman agents have been equipped with humanlike interactions for this very reason.

In addressing this investigation's first sub-question – on exploring the nature of the perceptive phenomenon – these coalescent factors should subsequently lead to a series of testable predictions for our interactions with home AI assistants – specifically that anthropomorphic schemas ought to be moderated by both extended contact with the agents and higher understanding or sense of control over their performance. Essentially, 'time' and 'understanding' should, therefore, ultimately dictate the strengths of users' humanlike schemas, ranging from none or very few anthropomorphic attributions to allocating extensive humanlike characteristics. These expectations formed the starting hypotheses for the first main experiment, on the Human (Figure 7).

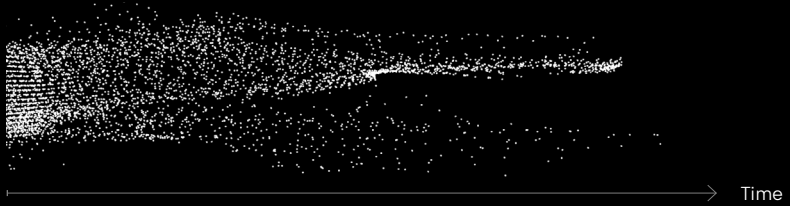
01 /
AGENT
KNOWLEDGE

Expected:
Decrease
over time



02 /
EFFECTANCE
MOTIVATION

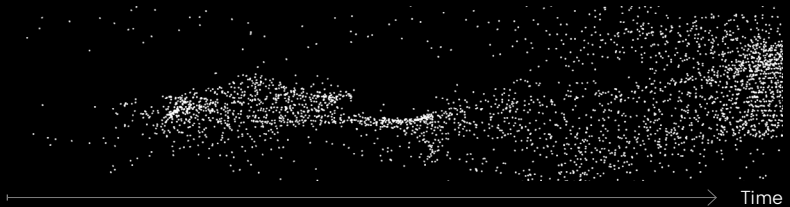
Expected:
Decrease
over time



7

03 /
SOCIALITY
MOTIVATION

Expected:
Increase
over time



[EXP01] THE LOOPHOLE

A. INTRO

The Three-Factor Theory of Anthropomorphism (Epley, Waytz and Cacioppo, 2007), in combination with Haslam's theory on dehumanisation (2006) (as a method for grading anthropomorphic attributions onto two dimensions of humanness – see Appendix E), has inspired a wide range of experiments investigating anthropomorphic representations in the design of AI, HRI and HCI (Złotowski, Strasser and Bartneck, 2014; Złotowski *et al.*, 2015). Much of the rhetoric on the benefits of personification in design is founded on this three-factor model, therefore, testing its predictions has repercussions not only for establishing the strengths of users' anthropomorphic beliefs and schemas but also for evaluating the arguments in favour of anthropomorphism within the field of design.

Hence, motivated by both the considerable application of this theory in design as well as the appeal of exploring whether the strength of humanlike schemas can be predicted by users' understanding and the extent to which they interact with their home AIs assistants, the first main experiment sought to investigate the correlation between *satisfaction*, *understanding*, *time* and the *nature of the anthropomorphic perception*.

The ensuing investigation followed a methodological approach based in Grounded Theory, with results subsequently subjected to statistical analysis.¹¹

¹¹ Considering, however, the overarching constructivist epistemology, I ought to furthermore state that no theory is, in fact, to be discovered, but instead constructed through an interplay between myself and the interaction with both the phenomenon and participants (Bryant, 2002). It has, however, become something of a shibboleth for researchers to claim to be constructivist as a matter of rhetoric only, and if I do make such a claim, its significance must be reflected upon. A constructivist epistemology has crucial implications for many research dimensions, but most importantly, it speaks to how its data should be viewed and interpreted. Therefore, at no point will there be a claim of 'objective truth' (regardless of the choice of method) but rather a version coloured by my own perspective, positions and values, acknowledged to the best of my ability. Thus, even the choice of tools – such as statistical analysis – is not justified by a presumed claim for 'objectivity', but instead by the demands of the many dimensions of the research tasks.

This simple statement, however, commands an indispensable explanation as to what substantiates this choice of method? Though undoubtedly an unorthodox rendition of design methods, the decision to pursue a formalised quantitative approach such as statistical analysis is justified by two distinct aspects: firstly, it sought to follow both social sciences and HCI research design conventions on the topic, in matching the methodological devices utilised to support this theory, and secondly, it was further reinforced by the unexpectedly large number of participants that ended up subscribing to the experiment (940).¹² Given such a large dataset, extrapolating information from open-ended questions would have been highly unrealistic. This consideration further impacted a set of open-ended questions that were included in a preceding version of the survey (in which no more than 60 participants were predicted to partake), which ultimately had to be compressed into a single question that could pass through a world-cloud analysis.

B. METHOD

I. APPARATUS

The experiment was conducted in the form of a survey developed on the Survey Monkey¹³ platform (2018) and posted on several specialised ‘subreddits’¹⁴ (discussion forums) on home AI assistants. The survey – comprised of ten questions divided into three sections – was aimed at people who owned one or

12 This is also done to avoid interpretation with regards to the choice of methods.

13 Survey Monkey is an online survey software that provides free customisable surveys including access to sample selection, bias elimination and numerous other data analysis tools.

14 “Reddit (launched in 2005) is a social news aggregation, web content rating and discussion website, ranked as #6 most visited website in the world with 234 million unique users (as of February 2018). [...] Registered users submit posts that contain a title, an external link or a self-written piece of content, which immediately become available to the whole audience of Reddit for voting and commenting. [...] The huge posting space of Reddit is divided into subreddits – self-created communities of users, united by a certain topic. Each subreddit and Reddit itself has a so-called “top page” (Medvedev, Lambiotte and Delvenne, 2019).

multiple personal digital assistants/home AI assistants, such as Amazon Echo, Google Home, Sonos One or Invoke.¹⁵ Prior to commencing the survey, participants were faced with an initial page containing a summary of the project, as well as a downloaded participant information sheet covering the legally required data protection information.¹⁶

II. MEASUREMENTS

As already stated, the questionnaire was divided into three distinctive sections, each addressing different aspects of the experiment: extended use and applications; satisfaction and understanding; anthropomorphic traits (see Appendix C). Participants were required to complete each section chronologically to ensure that possible biases, such as demand characteristics¹⁷, were thoroughly contained or avoided altogether. This necessity was driven by the impression that permitting participants to familiarise themselves with the last set of tasks – on anthropomorphic inferences – would have possibly influenced their answers to the previous sections.

The first section was aimed at primarily categorising the types of home AI assistants, the duration of time participants had them for, as well as the extent and nature (the functions employed the most) of their interactions with their devices. This section was designed to review the elicit agent knowledge hypothesis – that *time* and *extended usage* would curve anthropomorphic beliefs, ultimately resulting in users preferring object-oriented (such as technological) schemas as an anchor for induction over explicit humanlike schemas. Participants evaluated their responses on a Likert scale from 1 to 5.

15 At the time of the experiment, Apple’s HomePod was yet to be launched; nevertheless, a sizable number of participants chose to refer to their iOS’ Siri when filled in the survey (which is indeed the same AI that now powers the HomePod) – and therefore, findings are expected to perhaps equally capture users’ representations of Apple’s proprietary assistant.

16 Likewise, on the final page, the “submit” button had been prefaced by a statement reminding participants that by clicking the button they consent to participate in this study. For security reasons, data has been anonymised and stored on RCA’s cloud storage.

17 A situation in which the outcomes of the survey would have been biased by the experimenter’s expectations, creating a demand for participants to provide particular responses.

The second section was first intended to assess participants' satisfaction with their home AI assistants, similarly on a Likert scale from 1 to 5. Secondly, it sought to evaluate users' understanding of the backend technology (or algorithms) on a scale from 1 to 10. These questions were thus planned to broadly estimate effectance motivation's power to predict the extent of anthropomorphic attributions (which is predicted to be high in cases where users do not understand the behaviour of the nonhuman agent). In this vein, it was also necessary to test the strength of the frequently invoked rapport between anthropomorphism and user-satisfaction, which, in itself, would also indicate a degree of frustration.

The third and final section was designed to assess the range of users' anthropomorphic attributions, using a variation on a frequently employed scale, originally developed by Haslam in *Dehumanization: An integrative review* (2006). The variation of the scale employed in this experiment was derived from Złotowski's (2014) review of anthropomorphic traits (see Appendix E), in which anthropomorphism is mapped onto two distinct magnitudes of humanness: the Uniquely Human (UH) dimension, signifying intelligence, intentionality and secondary emotions, and the Human Nature (HN) dimension, represented by primary emotions, sociability and warmth. Lastly, this section also required participants to succinctly summarise their relationship with their selected digital assistant and to further describe their AI in brief.

III. PARTICIPANTS

As the link to the survey received a surprising number of up-votes (and was thus pinned to the top of the page), a staggering number of participants enrolled in just a couple of hours. The survey had to ultimately be closed within less than five hours from when it was advertised, firstly because statistical significance began plateauing and secondly due to the Survey Monkey platform slowing down due to the mounting contributions.

Therefore, a total number of 940 anonymous applicants from various cultural and social backgrounds participated in this survey, out of which a total of 874 (which were thoroughly and fittingly completed) were validated for statistical analysis. One evident bias resulted from the uneven engagement of males and females (89.6% male and 9.34% female), which, most likely, either reflects the demographic distribution commonly engaged in these specialised discussion forums or a general unevenness in consumers' interest with this technology (see Appendix D).

Secondly, another more elusive participant disparity was expected to bias results through the type of participation, which echoed perhaps a narrow category of users – early adopters/technology enthusiasts. As this was expected before commencing the experiment, different methods of recruitment were also employed, such as posting the survey on the Royal College of Art’s internal forum or various social media platforms, seeking to appeal to a broader audience. Nonetheless, participation was nowhere near as significant to counteract this unevenness and therefore had to, ultimately, be taken into account when analysing the results.

Lastly, it should be noted that since additional ethnographic/demographic data has not been collected, this representative sample only accurately reflects the characteristics of the Reddit ethnographic/demographic engagement and not of the broader population. This represents a limitation moving forward and additional research is needed to account for this potential bias.

IV. DATA ANALYSIS

As first specified in the three-factor-theory, when people lack adequate mental models, they will most likely use humanlike inferences as anchors of inductive reasoning. The elicit agent knowledge determinant thus predicts that, as we attain more appropriate knowledge structures, anthropomorphic beliefs would be less likely to manifest. This is expected to occur in any situation that gives users ample space to form more relevant mental models – implying that time and extent of use must weaken the attribution of humanlike traits.

Subsequently, it was predicted that a negative correlation would be found between the time participants have possessed an object with AI features or the extent of how often they used it and the extent of human qualities they assigned to it; yet, plotting the data into IBM’s SPSS¹⁸ software package and performing a series of bivariate statistics exposed the absence of any correspondence. This indicated that the degree to which people assigned humanlike traits *did not vary*, in any direction, *over time* and that the possible acquisition of new mental models had very little to no influence on anthropomorphic representations (Figure 8).

18 SPSS Statistics – name which originally stood for Statistical Package for the Social Sciences – is a popular software package used for interactive, or batched, statistical analysis. (SPSS Software | IBM, 2019).

While these findings appear to be pointing either towards a refutation or a clear limitation of the elicited agent knowledge theory within this context, alternative interpretations will be considered in the ‘Discussion’ section.

Secondly, the effectance motivation determinant forecasts that the more abstruse, misunderstood, or expectation-defiant an agent appears to be, the stronger the anthropomorphic inferences associated with it (Waytz *et al.*, 2010). Subsequently, it was hypothesised to find a definitive negative correlation between participants’ understanding of the backend technology and the extent to which they projected anthropomorphic traits onto their AI assistants. Except, precisely the opposite was found – a one-way ANOVA with *understanding* and *anthropomorphism* indicated with high statistical significance ($p=0.015$, with 0.05 being the minimum requirement for statistical significance) that the more participants claimed to be accustomed to the backend algorithms and their tasks, the more humanlike qualities they ultimately assigned to it, presenting yet another limitation to the effectance motivation thesis (Table 1 & Figure 9).

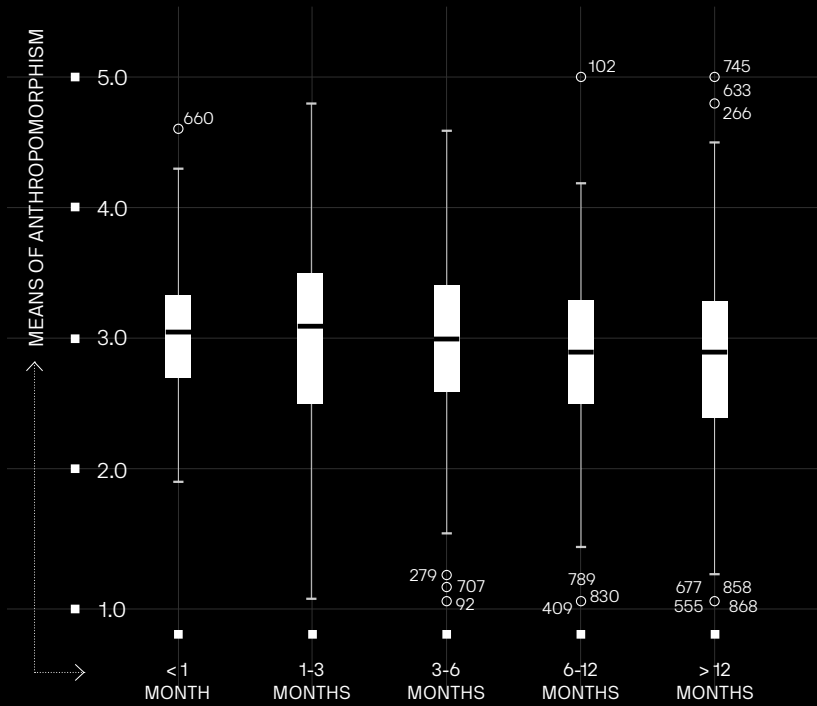
Anthropomorphism

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	6.994	4	1.749	3.096	.015
Within Groups	457.439	810	.565		
Total	464.433	814			

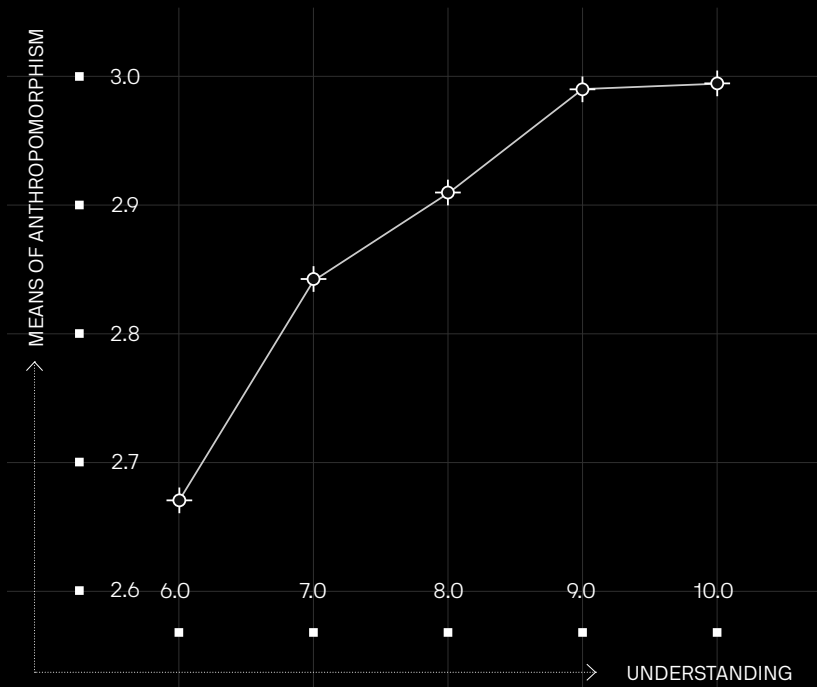
T1

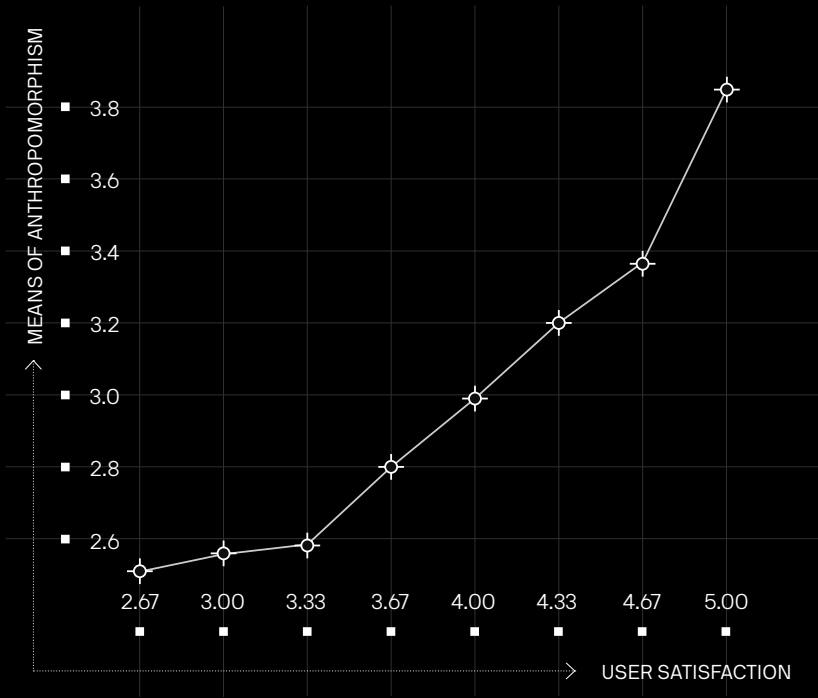
Moreover, the prevalence of the intentional stance in design, while rooted in the effectance motivation factor, is primarily justified by a hypothesised positive correlation between user-satisfaction and anthropomorphism. The theory, therefore, argues that personified designs activate readily available models of interaction, which, in turn, would increase users’ satisfaction with the product. However, a one-way ANOVA with *satisfaction* and *anthropomorphism*, while indicating with high statistical significance ($p<0.001$) that the extent to which participants assigned humanlike traits to their AI assistants was directly associated with the level of their fulfilment with the product, the relationship appeared to be correlational rather than a causal one (Table 2, Table 3 & Figure 10).

8

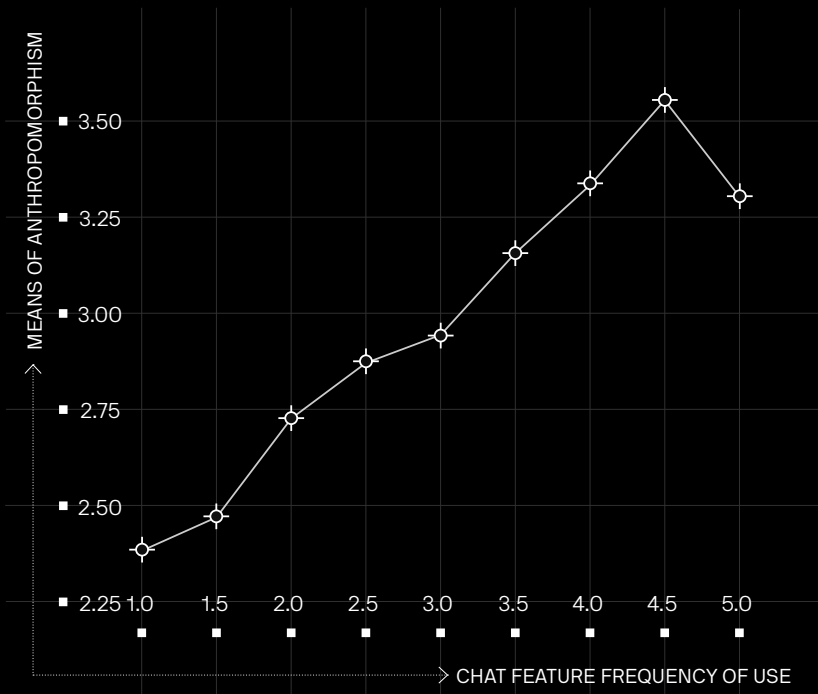


9





10



11

Anthropomorphism				
	Levene Statistic	df1	df2	Sig.
T2	.895	7	832	.510

Anthropomorphism						
		Sum of Squares	df	Mean Square	F	Sig.
	Between Groups	57.618	7	8.231	16.208	.000
T3	Within Groups	422.531	832	.508		
	Total	480.149	839			

The justification for this comes from performing an opposite ANOVA – on anthropomorphic qualities resulting in a higher degree of satisfaction – which was statistically insignificant. This finding proved to be a crucial element in revealing a perhaps more compelling explanation for the limitations of these findings, which will also be explored in the ‘Discussion’ section.

Lastly, while attempting to look for alternative correlations between usage and anthropomorphism, one noteworthy finding seemed to hint at the validity of the sociability motivation factor. A one-way ANOVA with *chat feature* (which measured how often people arbitrarily conversed with their AI assistant) and *anthropomorphism* indicated, with high statistical significance ($p < 0.001$), that the frequency with which participants engaged in dialogue with their devices informed how much they subsequently anthropomorphised, but not the other way around (Table 4, Table 5 & Figure 11). This seems to be an endorsement of the sociability motivation hypothesis – which states that the need for social connection would result in higher anthropomorphic inferences. Interestingly, therefore, people openly chatted to their devices because of their own psychological motivations, perhaps motivated by a need for sociability, and not because of the device’s perceived degree of anthropomorphism. Personifying designs could thus alleviate users’ need for social belonging, but the degree of personification appears ultimately to have limited to no significant influence on the amount of social pain it can mitigate (which seems to only be directly influenced by users’ own psychological determinants).

T4 Levene's test must be non-significant ($p > 0.05$)

T5 One-way ANOVA with anthropomorphism and chat feature

01

Anthropomorphism

Levene Statistic	df1	df2	Sig.
1.013	8	865	.424

T4

Anthropomorphism

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	40.465	8	5.058	9.583	.000
Within Groups	456.550	865	.528		
Total	497.015	873			

T5

C. FINDINGS

These statistical results, which seem to have exposed strong limitations for the three-factor-theory, invite a pressing need for scrutiny. In the first instance – while testing the elicit agent knowledge hypothesis – we found that the degree to which people assigned humanlike traits did not vary in any direction over time and that the possible acquisition of new mental models had very little to no influence on anthropomorphic representations. These results seem, at first glance, to come in direct contradiction with the three-factor-theory, which states that as users attain new information about nonhuman agents, their schemas will shift towards lesser anthropomorphic beliefs.

A closer inspection of the statistical results (Figure 12) might, however, reveal another mechanism at play, which links back to the caveat mentioned in the previous section: that anthropomorphic schemas may be extremely challenging to correct, especially in contexts where the attainment of further knowledge is likely to be inaccessible or obfuscated by the nature of the technology ('blackbox'). Consequently, the insight into the minimal alleviation of anthropomorphic inferences over time could be interpreted in two ways: either users do indeed acquire new knowledge structures, yet anthropomorphic beliefs remain unchanged (which would directly refute the three-factor-theory) or the

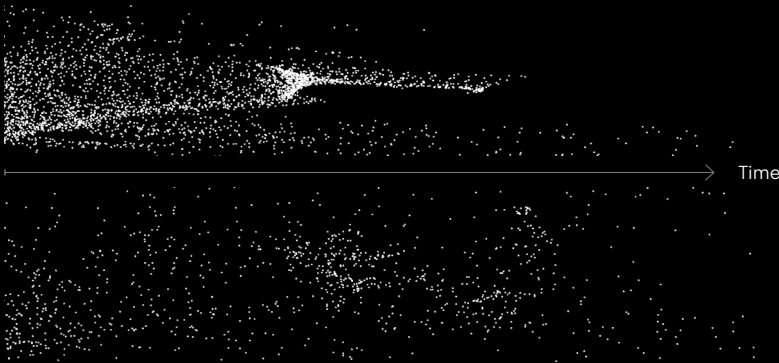
very process of correcting these schemas is obstructed by the obtuse inner workings of the tested AIs. The latter does not contest the three-factor-theory, but does, conversely, show a discernible limitation of the ‘elicit agent knowledge’ hypothesis, in this context.

By reviewing the authors’ own explanation for this possible limitation, we are indeed compelled to believe that it is the latter explanation that substantiates the results: “The inner workings of most modern technological agents are every bit as obtuse as the mental states of biological agents, but the incentives for understanding and effectively interacting with such agents are very high. This effectance motivation coupled with a general lack of understanding means that the tendency to anthropomorphize the workings of many nonhuman agents may be especially high” (Epley, Waytz and Cacioppo, 2007).

Nonetheless, these imprecise results reveal more than a mere limitation of the elicit agent knowledge hypothesis in the context of home AI assistants. On the contrary, they pose perhaps far greater implications for the field of design than if the initial hypothesis was refuted altogether – personified designs activate highly anthropomorphic perceptions that are especially difficult to correct. These schemas come with a clear set of expectations that, when defied, give rise to clear cognitive dissonance. The insight that, even with ample space for correction, participants’ unreliable humanlike mental models are carried over time with little to no correction exposes, therefore, an underlying difficulty in adjusting initial anthropomorphic projections to accommodate the AI’s expectations-defiant behaviour. Hence, any tension between participants’ expectations and the reality of the interaction would not be alleviated on its own, through extended use since users’ anthropomorphic beliefs will carry on, unmoderated, through time. This ultimately entails the need for designers to account for these dissonances ‘upstream’ in the development of personified interactions, instead of anticipating that tensions would naturally disperse through extended use. This finding accentuates a decisive shift in accountability, as designers become fully liable for mitigating potential arising tensions as part of the design process.

However seemingly trivial this finding might appear, it certainly reinforces the field’s overall obligation to develop systems that help users acquire more consistent and concrete mental models.

01 / AGENT KNOWLEDGE



E

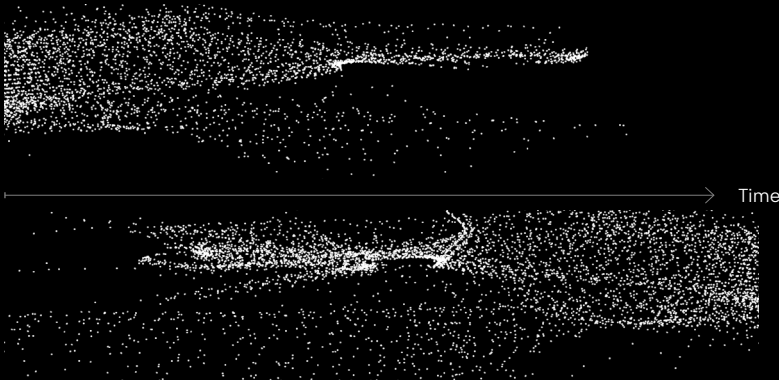
Expected: Decrease over time

Time

R

Result: No Correlation

02 / EFFECTANCE MOTIVATION



E

Expected: Decrease over time

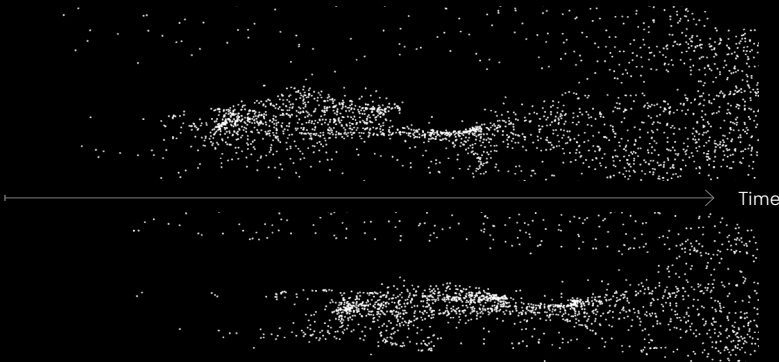
Time

R

Result: Increase over time

12

03 / SOCIALITY MOTIVATION



E

Expected: Increase over time

Time

R

Result: Increase over time

In the second part of the experiment – while testing the effectance motivation hypothesis – we found with high statistical significance ($p=0.015$) that the more participants claimed to be accustomed to the backend algorithms and their tasks, the stronger their anthropomorphic attributions, which once again stands in opposition to the tested theory. However, as we have seen in the previous analysis, we should not be so quick to dismiss the theory's validity. Instead, we ought to commence by looking for any plausible alternative explanations.

One such explanation could arise from a bias ingrained in the nature of the participants – who, as previously stated, likely included a disproportionate number of technology enthusiasts and potential hobbyist programmers. This category of participants would perhaps be more inclined to overstate their personal understanding of the technology, maybe referring to a more 'no-nonsense' understanding of composing 'skills' and integrating various IoT systems, rather than alluding to an understanding of more granular elements such as the NLP/ML¹⁹ algorithms that key to the functioning of these devices. As participants' understanding was only evaluated on a subjective scale, the representativeness of their responses could perhaps be quite different to what this scale initially intended to measure. It would, however, be expected that irrespective of the fairness of their perceived understanding, the scale should reliably reflect a similar degree of confidence and need for control, which ought to ultimately lead to fairly similar results.

Another possible explanation could further support the previous argument: the predisposition to anthropomorphise these nonhuman agents might be unusually high, perhaps even in the face of moderate understanding. This could be motivated by the highly personified nature of the interaction, which, regardless of the presence of alternative models, seems likely to prime strong anthropomorphic beliefs.

¹⁹ Natural language processing (NLP) is a subfield of computer science and artificial intelligence involving the study of natural language and its various computational aspects. NLP research focuses on understanding natural language communication with computers, extracting meaning from text, and generating natural language text (Nadkarni, Ohno-Machado and Chapman, 2011).

Machine learning (ML) is a subset of artificial intelligence research that enables computers to learn from data, without being explicitly programmed to do so.

Lastly, users' 'subjective understanding' and expectation defiant behaviour or ambiguity are not mutually exclusive. A percentage of users could be comfortable with vague behaviours and ambiguous responses, all while acknowledging these uncertainties as being part of the experience, which would simultaneously result in both high effectance motivation and high 'subjective understanding'. This hypothesis aligns with the finding that certain users are more capable of suspending their disbelief²⁰ for the sake of entertainment (Duffy and Zawieska, 2012).

Nevertheless, the aforementioned three alternative explanations could, at most, account for a non-correlation between participants' understanding and anthropomorphism, leaving the detected statistical correlation in the diametrically opposite direction decidedly unjustified. We must, therefore, conclude that the above rationalisations cannot adequately substantiate the evidence indicating that participants statistically *anthropomorphised more, not less*, the higher their reported understanding.

A development in interpreting the analysed data may result from coupling the above results with outcomes obtained from analysing the correlation between *user-satisfaction* and the attribution of *anthropomorphic traits*. Such an analysis appears to hint at a very different phenomenon. The justification for testing this in the first place comes from the field's attachment to the 'intentional stance' – which renders anthropomorphism as a design device, intended to instil users with a sense of competence in their interaction with humanised AIs. Indeed, AI design seems to benefit from attempts to elicit anthropomorphism. Such an effect is portrayed, for instance, by Kiesler and Goetz's (2001) study revealing that people were more likely to cooperate with robots that appear to exhibit playful behaviour, which were rated as being more extroverted than with robots that seemed more serious and therefore introverted. Experiments such as this appear to reveal that eliciting stronger anthropomorphic beliefs could lead users to experience an increase in engagement and subjective understanding, which could ultimately lead to an increased contentment with the product. Increased satisfaction by way

20 Suspension of disbelief: a "willingness to suspend one's critical faculties and believe something surreal; sacrifice of realism and logic for the sake of enjoyment". Most commonly employed in fictional storytelling (Duffy and Zawieska, 2012).

of apparently increased humanlike traits should, therefore, result from a lack of frustration and adversity with the product's performance or behaviour.

However, the results of this analysis seem to uncover a distinctive phenomenon, hinting at people's need for internal consistency. While *user satisfaction* and *anthropomorphism* are indeed decidedly correlated ($p < 0.001$), the opposite (between *anthropomorphism* and *satisfaction*) does not seem to be true. This would indicate that it may be user satisfaction that predominantly leads to higher anthropomorphic attributions rather than higher humanlike schemas resulting in increased satisfaction. This seeming challenge to the field's devotion to the intentional stance should, however, be taken with a pinch of salt.

These results appear to reinforce people's requirement for internal consistency, order or agreement among various cognitions and their 'relationship' with their AI assistant. This relationship is, in itself, an indication of already complex humanlike schemas (such as stereotyping). The intentional stance may still "serve as [an] effective method for improving the usefulness of certain agents" (Dennett, 1998) and increase user satisfaction and cooperation when combined with less elevated levels of anthropomorphism. Therefore, these results do not dispute the overall arguments in favour of pursuing anthropomorphism in design, though they do uncover a possibly more straightforward and pragmatic explanation than the one previously proposed. It seems that when highly personified designs elicit elevated levels of anthropomorphism, observed data indicates a need in people to solve internally inconsistent beliefs. This is evidenced by the fact that an increase in both users' satisfaction and understanding was correlated with a linear increase in anthropomorphic attributions, most probably as a means of commendation, conserving, therefore, a consistency among users' attitudes towards the product as well as their relationship with it. This observed phenomenon is then better explained by cognitive consistency theories – "a group of attitude theories, which emphasise people's need to maintain internal consistency, order and agreement among their various cognitions" (Hogg and Vaughan, 2017) – rather than an account of users' psychological predispositions. In this vein, participants ultimately appeared to have anthropomorphised more when, in fact, they were positively sanctioning (socially praising) agents' behaviour in order to alleviate a possible dissonance between their declared satisfaction or understanding and their evaluation of their AI assistant's humanlike traits (perceived as positive grading). These findings are further supported by a word-cloud analysis (Figure 13) of participants' short descriptions of their AI assistants, in which people who



declared a higher understanding, employed gentler and slightly more personified terms such as “smart”, “convenient” or “useful”, whereas a lack of understanding or below-average satisfaction was more frequently associated with less anthropomorphic, more objectifying terms such as “tool” “frustrating” or “limited” (see detailed breakdown in Appendix D).

This phenomenon appears, however, to be limited to already strong humanlike schemas, as we would typically not expect to find any need for consistency between users’ understanding or satisfaction with a non-personified object and a rejection of its humanlike qualities. Thus, in this case, the anthropomorphic scale no longer involves plotting a gradient of anthropomorphic beliefs; it instead echoes people’s attitude structures when it comes to already highly humanlike schemas. These results, consequently, appear to hint at a more straightforward explanation of an apparently complex phenomenon. Within this context, at least, users’ schemas of highly humanlike interactions could be inferred from studying their attitudes through a cognitive consistency lens (or cognitive dissonance lens). Such an approach will form the basis of a new investigation in the following chapters.

DISCUSSION

DELEGATION OF ACCOUNTABILITY

Let’s focus first on the borrowing of the three-factor theory of anthropomorphism within design and HCI practices. As noted, Epley *et al.*’s theory (2007) is first and foremost a cognitive account of anthropomorphism, localising the perceptive gradient not on the external humanlike traits that inform these predilections but on personal, psychological determinants. Failing to properly recognise this in design practice might lead practitioners to regard anthropomorphism as an inherently stable cognitive effect.

Thus, while advocates of eliciting anthropomorphism appear to be commonly compelled both by users’ compulsion for predictability and effective interactions as well as their necessity for social belonging, these justifications seem to highlight foremost a delegation of accountability to users’ own psychological motivations. The three-factor-theory is certainly an invaluable device for

measuring our psychological predilections, nevertheless, within the context of home AI assistants, users are arguably far outside a zone of predisposition towards anthropomorphism, and deep within a zone of already applying highly humanlike schemas. Within the presumed anthropomorphic gradient, this entails that anthropomorphism cannot be extracted from human cognitive predispositions external to the interaction. Anthropomorphism is a function of design. AI designers must, therefore, cater to users' social behaviours and dedicate further research towards understanding these behaviours and social mental models, rather than relying on people's intrinsic predilection for humanlike inductions and a presumed natural process of cognitive correction.

As we have so far seen in this large-scale experiment, anthropomorphic mental models, as proposed by the examined theory itself, are incredibly difficult to correct and are further sustained by highly humanlike interactions paired with a profound technological obscurity. This argument is substantiated in the first part of the experiment by the discovery that, even with ample space for correction, users' unreliable humanlike mental models are carried over time with little to no correction. The evidence suggests that any tensions between users' schemas and the nature of the interaction would not be alleviated on their own, and designers of personified forms must account for these dissonances 'upstream' in the design process, instead of so easily resorting to 'indistinguishability' as an incentive. Anthropomorphism may be inevitable, but it is not unproblematic.

Furthermore, while AI design seems to profit from eliciting anthropomorphic beliefs, there is quite clearly no distinct threshold upheld regarding the strength of these beliefs, leading to an unwritten conviction that the more humanlike an interaction appears to be, the higher the ease with which users will interact with it, eventually leading to an increase in user satisfaction. This experiment has shown, however, that this correlation of events is not so straightforward, and users' attitudes are perhaps better justified by more parsimonious explanations. Participants appeared to assign humanlike qualities as a means of maintaining an internal consistency in both their attitudes to, and in relationship with, their AI assistant. A state of mind in which users would have declared their satisfaction or understanding with the AI's performance, while concomitantly depriving it of admirable (humanlike) traits, would have been cognitively bothersome. This disharmony, known as cognitive dissonance, can perhaps be seen as participants' principal motivator in changing one or more contradictory beliefs, so that their attitude towards the AI, as a whole, would be in harmony.

However, behind the simplicity of this explanation lies a more considerable implication – these different cognitions would not have been disharmonious were the mental models of these agents not already highly anthropomorphic in nature.

COGNITIVE CONSISTENCY

Thus far, there has been no clear refutation of the three-factor-theory, but rather a qualification has been proposed regarding its ability to predict anthropomorphism in relation to highly humanlike designs. As has been noted, the three-factor-theory is only intended to describe a gradient of people's predilection for anthropomorphic inferences, as motivated by their own psychological determinants. It, therefore, cannot account for the relationship between *anthropomorphic features* and *anthropomorphic perception*. As a function of personification, users already employ highly anthropomorphic models, and psychological motivators do very little to account for the observed social behaviour. As such, it is within the remit of this study to consider alternative theories that may better suit unpacking the attitudes and beliefs that result in what appears to be an application of *cognitive consistency attitudes* in response to *humanlike cognitions*, rather than a gradient of anthropomorphic predilection. Hence, theories on cognitive consistency, with their focus on beliefs as the building blocks of attitude structures, have proven to be a more suitable device in the interpretation of this case study. Some of the most well-known theories, such as cognitive dissonance theory (Festinger, 1957) or Fritz Heider's (1958) balance theory – though different in how they outline consistency and inconsistency – are ultimately typified by people's tendency to find differences between their internal mental models and their personally observed external behaviours, further prompting an inclination towards consistency in this respect. In this experiment, we witnessed its effects on people's attitudes towards their AI assistants, which, in concert with the weight of evidence to be given to attitude change, could provide a solid ground for examining shifts and opportunities in users' anthropomorphic perceptions in the following chapters.²¹

However, considering the limited amount of research on anthropomorphic schemas per se, and the absence of theories on consistency or inconsistency in anthropomorphic beliefs, throughout the remainder of this study, the supporting

²¹ This, of course, involves a postulation that users' schemas are already significantly humanlike, which indeed appears to be corroborated by both practice and theory so far.

arguments that follow are not based on evidence of how other people reason about humanlike agents, but instead consider research based on how people think about other people. The support for this pursuit comes from psychologists and sociologists such as Epley, Wayts and Cacioppo (2007), who reason that “anthropomorphism itself involves a generalization from humans to nonhuman agents through a process of induction, and the same mental processes involved in thinking about other humans should also govern how people think about nonhuman agents.” This is further reinforced by research that establishes that the same neural system involved when thinking about other people’s activates when we make anthropomorphic inferences about nonhuman agents, and that the very same neural deficiencies that lead to inaccurate or incomplete attributions of mental states or emotions to people (such as in the case of individuals with Autism Spectrum Disorder), also affect reasoning about humanlike agents (Castelli *et al.*, 2000; Iacoboni *et al.*, 2004).

To conclude, the anthropomorphic perception is not at all fundamentally “observer-dependent”, therefore giving weight to the opposite of this investigation’s hypothesis. Indeed, it is strongly dependent on the choice of anthropomorphic affordances consigned to the artefact. With anthropomorphism being a function of personification, the hypothesis of this investigation shifts to the equally important aspects of design and how they impact expectations and the formation of anthropomorphic mental models. Consequently, let’s consider this social AI in the following chapter.

02

THE AI

As was outlined in the previous chapter, the first experiment indicated a highly persistent humanlike process of induction, cued, most likely, by the particularly personified designs of AI assistants and further self-sustained by the entrenched obscurity of their technological inner workings. In my contention, it is precisely the complexity of this mechanism, paired with AI design's attachment to the concept of technological indistinguishability (a significant design incentive mentioned in Chapter 1) that poses challenges. Contradicting the initial hypothesis of anthropomorphism as a 'mindless' reaction, it has been noted that a more complex Human-AI dynamic exists, in which the users' inferences are inseparable from the interaction.¹ With the previous experiment exposing anthropomorphism as not entirely observer-dependent, but also as a function of the artefact, this section focuses on the AI, examining the relationship between the intentional or unintentional use of technological cues and matters of perception. This inseparable coupling of the environment and direct perception introduces the concept of *affordance* – a term that, despite its ambiguity, is incredibly popular in HCI design.

Thus, this chapter proceeds as follows: in the first part, I introduce classifications of affordances from both James Gibson's approach and that of Don Norman and elaborate on how the use of metaphors in design represents a means of communicating both usefulness and usability to the users. With the human metaphor being a design choice that gives an AI system a set of affordances, I hypothesise that the existence of *false affordances* (i.e., widely perceived and non-existent altogether) is very problematic. As a result, in the second part of this section, I am experimenting with the mechanism through which *anthropomorphic features* lead to *anthropomorphic expectations* in the same methodological approach as before: first by briefly consulting the literature on anthropomorphic affordances

¹ This clarification is foremost required as we often find a disagreement between the presence of social behaviour and anthropomorphism, sometimes presented as two distinct phenomena, especially in a now-classic experiment, widely known as The Media Equation. Reeves and Naas (1996), in an experiment done at Stanford University, established that people extend social behaviours to media regardless of how humanlike they appear to be: in one of their extensive studies on people's relationship to media, they asked participants to evaluate the performance of a computer in two different scenarios. In the first one, participants had to score the performance on the same computer they performed the given tasks, whereas the other group had to rate it on a computer different to the one they had utilised in the experiment. Those in the first group consistently gave higher ratings to the computer, unmistakably applying the same rules of politeness as they would in a typical social situation.

and subsequently by proceeding with the unfolding identification and the need for further experimental evidence.

However, reviewing the interaction between personification and anthropomorphism brings about further considerations, such as, on what basis do people make inferences about “an implied other”² and whether we would expect anthropomorphic attributions to coincide with the mechanism employed when making inferences about other people’s intentions, motives and behaviours. In short, studying the inseparability between the artefact and perception entails foremost an understanding of the interactions between humanlike design cues and perceived properties (which may or may not actually exist). Thus, the experimentation phase proceeds in two stages: first, by reviewing whether the perceived properties are built on users’ pre-existing human models and knowledge of other people (Just Like Me), and secondly, by probing whether they are identical or diverge with some respects to their human corresponding models (But Not Exactly).

AFFORDANCES <> ANTHROPOMORPHISM

The concept of affordance, while very popular within the HCI community, remains ambiguous, with widely varying uses of the term. The incongruity can be traced back to its two main definitions (Gibson’s and Norman’s) that, despite sharing some fundamental aspects, are not necessarily interchangeable entities. This difference essentially hinges on the nature of the relationship between the environment and the actor. Establishing which perspective this research adheres to is, therefore, needed to elucidate the role of the designer and how we draw the distinction between usefulness and usability when designing AI agents.

² Here, making an analogy between social psychology’s study of how people are affected by other people who are psychically present or who are imagined to be present or whose presence is implied (Hogg and Vaughan, 2017). However, in contrast to social psychology’s ‘implied human presence’, which refers to people actually conceiving of others, in the case of personified designs, ‘the implied other’, a clear anthropomorphic inference, is often catalogued as a ‘perceptual mistake’.

This section focuses on the distinguishing areas demarcating these different relationships.

Significantly, while both denote the action possibilities to an individual by an environment, in Gibsonian terms, affordances exist independently of the individual's ability to perceive them. Therefore, they are epistemologically objective in that their presence does not depend on the actor's meaning or interpretation. To James Gibson, the perceptive psychologist who initially coined this term in his seminal book *The Ecological Approach to Visual Perception* (2014), affordances are invariant to the needs and goals of the actor. To illustrate this, Gibson gives the example of a solid, opaque surface extending under one's feet. The combination of solid opaqueness specifies an affordance of walkability regardless of whether that is perceived or not by an actor. Let me be clear: the affordance of walkability exists as a property of the surface in question, yet the ability to perceive it may be dependent on the action possibilities of an actor. As such, the actor forms a frame of reference, but the affordance exists regardless of it, independent of the actor's experiences and culture.

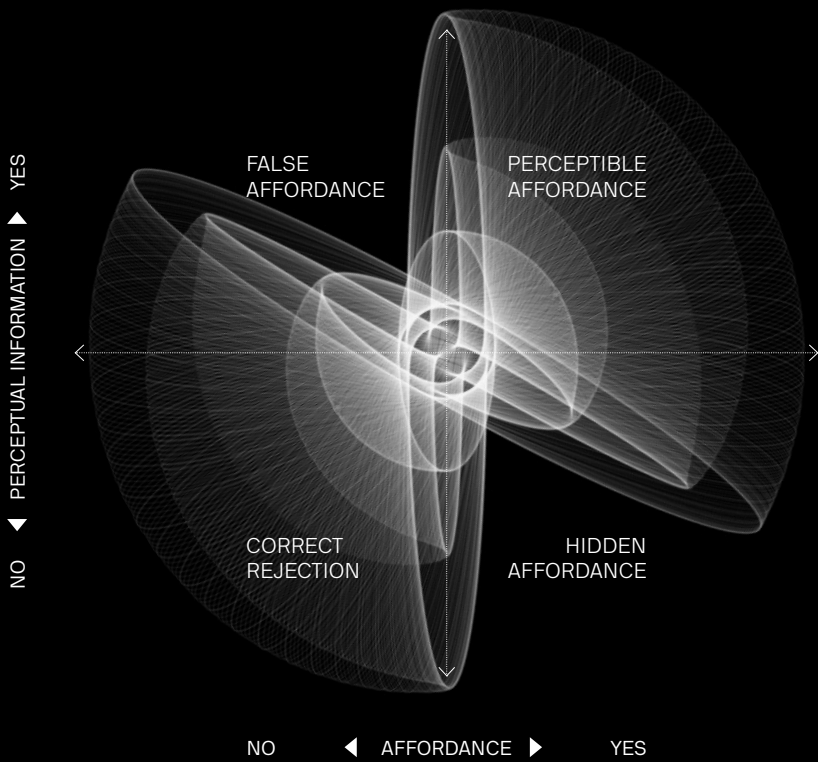
This definition can be contrasted with Norman's affordances, as popularised in his book *The Psychology of Everyday Things* (1988), whereby he appropriated the term and reinterpreted it to refer not only to the actual properties of an environment but also to the perceived ones (which may or may not be actual properties of an environment, but which represent affordance nonetheless). To some extent, we could say that Norman's affordances are epistemologically subjective yet still derived from a subject-object interaction. This distinction is made more evident from an endnote in Norman's book: "The notion of affordance and the insights it provides originated with J.J. Gibson, a psychologist interested in how people see the world. I believe that affordances result from the mental interpretation of things, based on our past knowledge and experience applied to our perception of the things about us. My view is somewhat in conflict with the views of many Gibsonian psychologists, but this internal debate within modern psychology is of little relevance here." (1988, p. 219). Of course, Gibson and Norman had two very different intentions, one being primarily interested in how we perceive an environment, the other with how we might manipulate the environment to signify a perception. Resulting from these divergent goals, Norman's view was that it is only what is recognised as an affordance that counts and that designers should forget about affordances in Gibsonian terms and focus on the signifiers (or perceptual

information) that lead to the perception.³

Therefore, these contingent views present two main demarcations: one is that in Gibson's claims, the existence of affordances is not dependent on values or meaning, while in Norman's view, they are tightly coupled with actor's past knowledge and experience. Secondly, in Norman's view, perceived properties may or may not actually exist, whereas to Gibson, they exist and are independent of interpretation. This difference forms the basis of this investigation's partial adherence to Norman's framework – and I say partial because, concurring with McGrenere and Ho's view (2000), I believe Norman's view to also be incomplete.

By favouring the importance of perceived affordances (as Norman describes them) as almost independent of an object's perceptual information, we are circling back to where this thesis started: an overemphasis on the role of the actor in this dynamic. It is precisely this difference that, in McGrenere and Ho's view, caused confusion in the HCI community. In his original definition, Norman collapsed two very important but different, and perhaps even independent, aspects of design: *designing the utility of an object and designing the way in which that utility is conveyed to the user of the object*. This confusion is more clearly illustrated by Gaver, who clearly draws the line between these two aspects of design: "Distinguishing affordances and the available information about them from their actual perception allows us to consider affordances as properties that can be designed and analyzed in their own terms" (1991, p. 81). As such, an affordance is not only dictated by the utility of an object (as in Gibson's view) or by the perceptive act (as in Norman's view) but also by design's choice of conveying that information to the actor. To address this, Gaver introduced a framework for separating affordances from the perceptible information available, demarcating them in 3 different flavours (Figure 14): *false affordances* (when the perceptual information is false), *perceptible affordances* (corresponding to Norman's perceived affordance, i.e. when there is information and the affordance is correctly perceived) and *hidden affordances* (when there is no perceptual information yet the affordance exist; which more closely aligns with a Gibsonian view).

3 This different viewpoint was shown simply in the now canonical example of affordances, namely the door handle example. In considering a door with no handle and no flat panel, it would be difficult for an actor to interpret the direction of opening. Following Gibson's definition, the simple fact that the door can be opened is sufficient to confer it that affordance and the door would need no further information to specify its existence. This stands in contrast to Norman's view in which the affordance does not exist in the absence of information to specify the action possibility, such as a handle to signal direction.



It should be noted that Gaver's definition meshes Norman's and Gibson's views in a somewhat contradictory manner in that some affordances can exist for the actor, yet they do not necessarily exist as part of the object. In other words, just because we perceive that something affords an action, it does not mean the environment indeed allows that action. To illustrate this, in Norman's view, the existence of a solid opaque surface underneath one's feet automatically grants it the affordance of walkability regardless of whether the surface in question is concrete or thin ice. In Gibson's view, the affordance belongs only to the former, whereas in Gaver's framework, the affordances exist in the actor's mind for both surfaces, yet the environment only affords it for one. Therefore, in this framework, what matters most is whether the perceptual information that specifies the affordance to the actor is false or not.

This framework is fundamental moving forward as I am hypothesising that with humanlike design, it is precisely this complexity of conveying perceptual information that leads to situations in which one *perceptible affordance* signifies new affordances that do not, in fact, exist (*false affordances*). In my contention, this lends to the challenges of leveraging anthropomorphism in design.

As with other systems of metaphors that give systems a set of affordances, the human trope was analogously envisioned to leverage users' past experiences and interactions with other people. While more metaphorical approaches to humanlike design have primarily relied on the abstracted functions of a conversation (such as feedback), the more literal tradition of designing home AI assistants has arguably placed an emphasis on simulating conversations as almost indistinguishable from their genuinely interpersonal equivalent – heavily relying on user's suspension of disbelief on whether these algorithms do indeed think, reason, observe and communicate like other people. Given this literal approach to the conversation metaphor, mapping what explicitly represents an affordance and how to convey the information about it is unsurprisingly more challenging than, for instance, exploring the affordances pertaining to the prevailing 'direct manipulation' metaphor of HCI (Fineman, 2004). Consequently, the more abundant sphere of salient cues that inform users' mental models and perceived affordances remains, understandably, only scarcely charted.

There does, however, seem to be an unceasing pressure to fill this sizeable gap in knowledge, which has led most studies on anthropomorphism in design or HCI to focus primarily on identifying and classifying this plethora of cues. Yet, isolated

studies on empathetic capabilities, cooperation, body movements, friendliness/seriousness, amongst others, perhaps only scratch the surface of the vastness of signifiers that prime users' schemas of 'an implied other'. To name a few, some of the earliest experiments on personifications have shown that people generally consider women's voices more agreeable than men's (Crowelly *et al.*, 2009), or that when the voice is uncannily 'natural', users tend to overestimate the capabilities of a system, which leads to a breakdown in communication (Moore, 2017). Conversely, in Human-Robot Interaction, studies such as Syrdal's *et al.* (2008) have demonstrated that a higher degree of personification is linked to higher expectations in terms of robots adhering to proxemics norms, or Bethel, Salomon and Murphy's study (2009) showing that participants were calmer when interacting with "non-anthropomorphic" robots "operating in an emotive mode versus a standard, non-emotive mode".

This research, however, seeks not to produce a map of such cues. This is firstly because, in my opinion, the act would needlessly overstate the role of design choices on anthropomorphism, just as much as studies on anthropomorphism frequently overlook design's use of perceptual information to prime anthropomorphism. Secondly, I am assuming that 'affordances' within a more literal humanlike interaction paradigm might be too numerous or too subtle to be considered effectively and in a non-reductionistic way. Thus, finding a way of isolating false affordances or those that produce misinformation, might prove a considerably more efficient way of tackling both the overarching research question as well as offering further insights on the practice of design that suggests these affordances.

In trying to acknowledge the inseparability of these two dimensions, I am proposing two experiments that reveal not the affordances, but rather, the fidelity of the mechanism behind the formation of the mental model and how perceptual information may give rise to *sequential affordances*.⁴ *In short, I am investigating the relationship between perceptual information and false affordances* – which is not to say that the affordances are false, but rather that the information is false – within anthropomorphic models.

4 According to Gaver, when one perceptible affordance leads to information indicating a new affordance, these can be denoted as sequential affordances – affordances that reveal themselves over time. These types of affordances should be expected in the humanlike metaphorical design framework whereby the presence of one perceptual information leads to many other inferences (Gaver, 1991, 1996).

[EXP02] JUST LIKE ME

A. METHOD

I. APPARATUS

The following experiment (intended as a continuation of Sau-lai Lee and Sara Kiesler's (2005) research on human mental models of robots) has been devised to investigate the mechanism through which users approximate agents' knowledge based on perceptible affordances. In Sau-lai Lee and Sara Kiesler's investigation, in which they required participants to estimate the likelihood of a robot produced in either China or the USA to recognise landmarks from either Hong Kong or New York, the authors contended that given minimal information, such as the robot's place of origin, participants readily employed mental models of other people to predict the robots' competencies.

However, since participants did not, in fact, interact with the robot, but rather briefly observed it perform with an experimenter, results are not conclusive as to how the participants justified these extrapolations and, therefore, as to how they constructed their mental models. They could have fashioned their mental models based on assumptions of the engineers that built the robot, rather than the system itself. This was clearly a false affordance, but whether it resulted from a perceptual design cue remains unsettled. In the authors' own words, "the data do not tell us how participants justified these extrapolations. Did they believe that the Hong Kong (or New York) engineers who built the robot also put information about tourist landmarks into a database accessible to the robot? Did they believe the robot in Hong Kong (or New York) had direct experience with landmarks?"

In this experiment, I am interested in exploring the mechanism behind users' mental model formations as resulting from the pickup of design information/misinformation. As such, it is trying not only to further probe the findings from the previous experiment on anthropomorphism as a cognitive theory (as opposed to a mindless reaction), but also to address the second sub-question of this study: *Are anthropomorphic expectations considerably different from their human-reality counterparts?*

To do so, this investigation, designed as a Wizard-of-Oz (WoZ)⁵ experiment, was intended to estimate the kinds of attributions primarily employed when forming mental models of new humanlike AI agents. The preferred mode of exploring the mechanism behind mental model formation entailed an indirect extrapolation of participants' estimations of the agent's presumed knowledge – in this case (and inspired by the above study), knowledge of various landmarks from around the world.

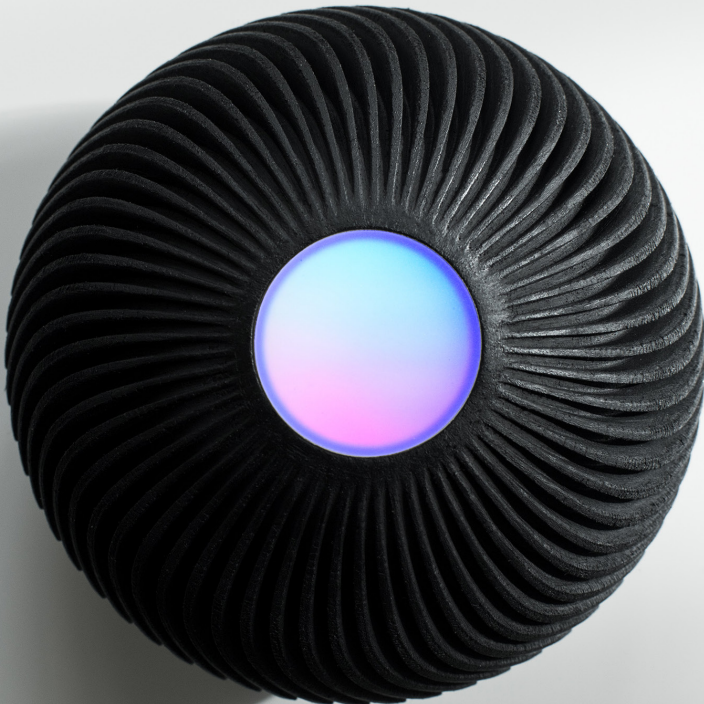
However, unlike the aforementioned study, this research required participants to directly interact with the AI agent by playing a guessing game modelled after the popular board game, *Activity*. Participants were thus invited to play the game with a novel Indian English (IndE) speaking female AI assistant “capable of analysing vast amounts of descriptive information” – in which they were asked to verbally describe a series of modern buildings from around the world while avoiding the use of eponymous terms. Using this approach with six distinctive landmarks (see Appendix F), one of which was located in India, the purpose was to detect whether the perceptual design information of an accent leads to sequentially false affordances (such as knowledge estimations).

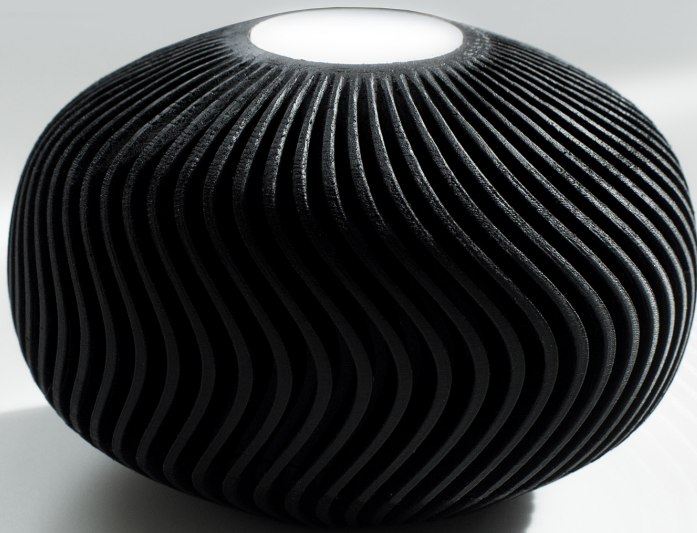
II. MATERIALS

To ensure that participants were indeed forming new mental models instead of relying on pre-existing ones, this experimentation entailed the design of both the physical and digital interaction. As a result, the physical device and the virtual interaction, respectively, were deliberately designed for this investigation based on the following criteria:

- + The form was designed to broadly resemble existing personal assistants to maintain conventions and avoid confusion over identity or function yet different enough to encourage the formation of new mental models. This was done to avoid ‘contaminating’ the experiment with any unforeseeable perceptible affordances. As such, the device was designed in a minimal fashion, featuring a generic black appearance in order to adhere to the sleek, minimally black design convention of personal assistants (see Appendix G).

⁵ The broad range of responses required by this study would have been too time consuming to replicate through traditional prototyping.





- + Following existing conventions on conversational feedback, the device featured an LED panel that provided rich, colourful feedback based on the performed task such as *awake, listening, processing, answering*, similarly striving for the uniformity and consistency accustomed users might expect from a virtual assistant (Figure 15 & Figure 16).
- + The agent was designed to convey the perceptual information of a distinctively recognisable accent to test the nature of the resulting perceptible affordances and whether it would sequentially lead to the hypothesised false affordances.
- + Similarly, the interface, though being operated by the experimenter (WoZ), had to nevertheless resemble current voice-activated interfaces. As such, the interaction was initiated by phrases such as “hey Veena”, and the IndE speaking female’s voice responses were pre-designed in a generic manner. The persona employed to deliver the lines was derived from one of Apple’s main OS system voices – Veena.
- + These considerations were taken to ensure that the only variable to be tested in this interaction would be the perceptible information of an accent.

The experiment was performed in a large study room with two desks and two chairs. On the desk at the top of the room, participants would find a laptop displaying the project brief and a test round of the game, as well as the device and a consent form (see Appendix N). The other desk, on the far side of the room, featured only a laptop from which the device’s responses were being operated by the experimenter.

III. PROCEDURE & MEASUREMENTS

Participants were seated at one of the desks, facing both the device and a laptop on which a slide presentation was listing a series of randomised modern buildings from the selected six different countries (Figure 17). They were then briefed about the apparent purpose of the experiment – *‘that of assessing the devices’ capacity of examining descriptive information – based on their untimed description of each building’*. Concealing the real intention of the experiment was a prerequisite in being able to measure the more elusive nature of perception reflected in the amount of time participants deemed necessary to describe each



of the six landmarks.⁶ As such, they were permitted as much time as they needed to describe each landmark before moving on to describing the next, and only after describing all the buildings, the AI ‘responded’ with its guesses. This was done to avoid the bias of the experimenter influencing the time required to describe a particular entry.

The buildings utilised in this investigation were therefore chosen based on the following criteria:

- + Semi-known – only by those who have either visited the country or possess a more extensive general knowledge of modern architecture. This measure was taken to lessen the influence of participants’ own extrapolation of the perceived difficulty of recognising each building, and therefore, in a control condition, spending roughly the same amount of time describing each entry.

⁶ It should be noted that the choice of a quantitative method as opposed to a more qualitative approach, such as interviews or emotion mapping, was dictated by the demands of the research task. The more elusive nature of anthropomorphic perception vanishes in a system of intentional self-perception. This is further supported by the results of the first experiment, in which a qualitative assessment of participants’ rationalisation of their assigned qualities to the AI agents revealed a ‘correct’ rejection of anthropomorphism.

- + Distinctive physical characteristics to ensure that each building could have easily been guessed in a classic game of Activity, if described correctly.
- + Buildings were chosen from five different control countries (China, UK, USA, Japan, Brazil) and the tested country (India) and were randomised for each participant (see Appendix F).
- + The Indian entry alternated between two different buildings, to contain the likely influence of any variation in the difficulty of describing the tested entry.

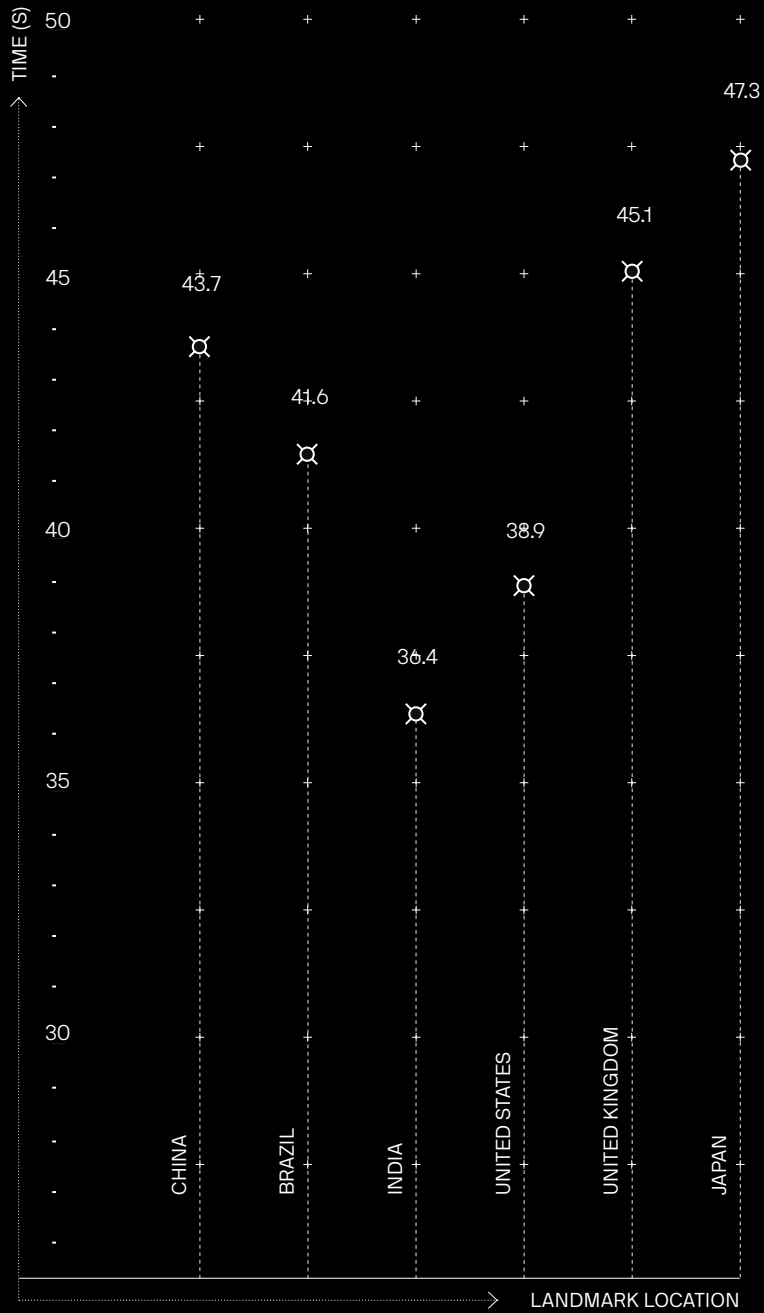
Before starting the experiment, subjects were allowed one test round to become familiarised with the rules of the game and begin the actual experiment from a more relaxed state. Informed consent was obtained after providing participants with ample information about the nature of the study and assessing their comprehension in a written consent form prior to initiating the research activity. Participants were also video-recorded when prior consent was given. For privacy reasons, data has been anonymised and stored on RCA's cloud storage platform (Google Drive). At the end of the game, subjects were informally asked about the device's performance as their overall (qualitative) reactions were recorded.

IV. PARTICIPANTS

A total of 18 participants (9 females, 9 males) from the Royal College of Art volunteered to participate in this experiment, by responding to a call posted on the University's forum. They were all fluent English speakers, and only 4 indicated that they have never played a game of Activity. None of the participants ever owned a personal digital assistant and all claimed never to have directly interacted with a physical one (though all confirmed an overall familiarity with virtual assistants such as Siri). Thus, the probability of them forming novel mental models was predicted to be high, although perhaps still heavily relying on certain pre-existing assumptions.

V. DATA ANALYSIS

By computing the amount of time each participant expended before moving on to the next entry, the data revealed that participants spent roughly 42 seconds on describing each building, with an average of 6.9 seconds less (16.4%) on describing the Indian landmark (Figure 18), in both control conditions (i.e.,



regardless of the complexity of the building). In fact, each participant spent the least amount of time describing the Indian entry out of all the six, regardless of the order in which it was being presented, hinting at a strong tendency to resort to subconscious expectations of the interaction. Their overall end reactions to the interaction were noted for more qualitative interpretations in the next section.

VI. RESULTS

To recap, participants saw six modern buildings from around the world, which they had to describe to an IndE speaking AI assistant. They were informed that the aim of the experiment was to *'assess the device's ability of analysing descriptive information'*, and that their performance in the game was at no point being evaluated. In reality, the only measured parameter was the amount of time participants spent describing each entry as well as their general reactions at the end of the experiment.

By unconsciously spending 16.4% less time on describing the Indian building, results strongly suggest that participants did, indeed, estimate the knowledge of the device based on perceptual design information such as accent and name. Thus, in assessing the nature of anthropomorphism, participants extrapolated from their own expectations of people with Indian accent as a means of assessing the extent of the devices' abilities, thus resulting in a series of false affordances (knowledge estimation based on place of 'origin'). This further implies that users undeniably employed the kinds of attributions they would typically engage when justifying other people's behaviours as an anchor for building their anthropomorphic mental models.

Additionally, as the interaction was controlled from another computer, in each case, the agent was instructed to deliver the correct responses to five out of six entries, alternating between providing a wrong answer to either one of the control buildings or the Indian entry. As anticipated, when questioned on the device's performance, in the interactions whereby the agent incorrectly guessed the Indian landmark, participants were largely more dissatisfied than those who received a wrong answer for any control buildings. This is most likely a result of the cognitive dissonance ensuing from a mismatch between expectations and reality, substantiating the existence of a false affordance. These final, qualitative measurements, however, must be taken with a pinch of salt – as participants were directly solicited

to evaluate the performance of the device, their responses were most likely biased by their interpretation of the experimenter's expectancies (demand characteristics) and may only expose an incomplete picture of participants' mental model.

B. FINDINGS

ANTHROPOMORPHISM RESULTS IN FALSE AFFORDANCES

In spending considerably less time describing the Indian landmark, participants were most likely extrapolating from their knowledge of other people when correlating an accent with a higher probability of identifying buildings from a place of origin and thus subconsciously providing less detailed explanations. In other words, subjects were unable to account for the more evident external limitations – such as the universality of the algorithmic processes and unimportance of the device's accent.

The extrapolation of specialised knowledge based on the most subtle signifiers (an accent), even in the face of overwhelming evidence against it, firstly suggests that anthropomorphic perception closely mimics the mechanism through which we make inferences of other people. As such, in the context of humanlike interactions, the presence of controlled perceptual information leads not only to correct perceptible affordances (such as perhaps the ability to 'understand' Indian English expressions) but also to nested apparent affordances that do not necessarily exist (knowledge of landmarks). This lumping of information most notably results in an inability to correctly reject these affordances, contributing to the argument against personification in design (as inescapably leading to the false expectations of erroneous mental models).

On the surface, it perhaps seems of less importance to scrutinise the perceptual mechanism behind this lumping of the information in the context of design, yet attempting to understand it may contribute to how designers convey utility through signifiers. Thus, to clarify these complex affordances, in my view, the lumping may be the result of a well-known mental heuristic – the fundamental attribution error – a cognitive bias referring to people's tendency to downplay situational explanations for an observed behaviour while overplaying person-

ality-based explanations.⁷ This, in turn, means that signifiers pertaining to the internal properties of an agent (such as personality) will be preferred over those pertaining to external (technical or form-based) properties, despite the obvious absence of any ‘internal’ traits.

These findings appear to predominantly point towards a hierarchy of relevant signifiers, in which case anthropomorphism arguably becomes partial to the perceptual information regarding dispositional traits (gender, name, skills, personality) – preferring these cues over other intended signals (such as shape, technical limitations, brand identity).

Accordingly, when it comes to designing humanlike interaction the indication that much of users’ interactions result in a lumping of affordances because of a mental heuristic most notably denotes a necessity to first attend to the more salient ‘internal’ signifiers. This work is, of course, at an early stage, and further research should address the equivalence between human personas and digital personas, whereby we might find that although certain expectations are seamlessly transferred, components such as factual knowledge may differ between the two (see Suu Kai’ study).⁸ This is further explored in the next experiment.

FALSE AFFORDANCES = COGNITIVE DISSONANCE

Following the discussion initiated in the previous chapter, the clash between users’ anthropomorphic expectations and the agent’s behaviour inevitably begets a series of cognitive inconsistencies. Significantly, these are the result

7 We rarely ever approach the task of ascribing causality for behaviours in an entirely objective, impartial manner, as the cognitive mechanisms responsible for attributions are themselves subject to imperfections. Heider (1958) was convinced that since people’s internal drives and intentions are entirely concealed from us, we only infer their existence once we have already exhausted all possible external causes. This was notably debunked, leading to what is now known as the *correspondence bias* or the *fundamental attribution error* – an effect which can be seen in the willingness with which we consider others’ actions and behaviours as a reflection of who they are, attributing behaviours to stable internal properties rather than the situations they are in.

8 Nevertheless, isolating user expectations in the absence of a studied interaction would only be based on assumptions from our knowledge of other people, inevitably disregarding the more nuanced beliefs of why an agent behaves the way it does.

of apparent affordances that suggest different actions than those for which the system has been designed, thus leading to false expectations and errors. While the mismatch leading to cognitive dissonance is not unexpected, what ends up becoming a false affordance is far less foreseeable. This is largely because research on the mechanism between personification and anthropomorphism is still poorly mapped and understood.

This experiment, however, has managed to isolate one significant factor contributing to users' expectation formation, reflected in their biased assumptions that the agent's actions depend on what 'kind of person' it portrays. Thus, users' expectations of a system are heavily influenced by their opinions of a certain persona (stereotyping – which in itself is a form of lumping information) – in our case, of an Indian-speaking persona. Significantly, this goes both ways.

REINFORCING STEREOTYPES

One noteworthy aspect arising from the above findings draws immediate attention to how design choices could (and often do) condition users to fall back upon antiquated perceptions of gender or ethnic stereotypes. It must first be noted that this conversation could easily form the basis of an entirely separate piece of research, and for this reason, in this section, I will only briefly mention the implications it bears to the design process. In fact, significant research has already addressed the more pressing long-term repercussions of employing a predominantly female-voiced persona as a digital assistant – such as United Nations' study *I'd blush if I could* (borrowing its heading from Siri's response to being told "Hey Siri, you're a bi****"). UNESCO's publication (2019), reflecting on the assistant's submissiveness in the face of gender abuse, is ultimately aimed at exposing the gender biases ingrained in this pervasive technology. This section is, thus, not intended as an extension on this discourse, but instead as a brief discussion on findings that seem to validate these already heavily addressed concerns.

Considerably, the fact that participants' attitudes towards the device were principally influenced by their expectations of a distinct persona – an Indian female assistant – is a powerful illustration of how projecting digital assistants as belonging to distinct social groups would inevitably reinforce potentially harmful stereotypes. The mere association between the device's function and its

persona unavoidably invites the plethora of expectations associated with both the function and the 'type of person' it portrays, and thus, regardless of whether it meets users' expectations or not, it still conditions people to fall back upon pre-existing schemas. Ultimately, while this experimentation only investigates knowledge estimation when presented with cues of a particular nationality, it inescapably implies the existence of many other associated assumptions and their implications for users' mental models.

While many studies advocate for "proper safeguards against hostile, abusive, and gendered language" as means of moving away from reinforcing biases, whether it is even feasible to employ a distinct persona without inviting the preconceptions associated with it remains largely unanswered. We might argue that the mere priming of certain expectations inevitably strengthens their existence, as their nature is ultimately culturally derived and historically situated.

This, of course, has worrying implications for design's pursuit of attaining increasingly humanlike systems rooted in indistinguishability, since design choices might ultimately have overwhelming cultural and sociological implications on the way in which we interact with other people in general. Finding alternative, more accurate signifiers, therefore, represents a research necessity. The aim, eventually, should be to find solutions devoid of cues of gender, ethnicity, nationality or origin in establishing a best practice.

As a result of these identifications, in the next experiment, I set the goal of exploring how a less-literal rendition of the human metaphor influences anthropomorphic expectations and whether, based on the context of the interaction and on cueing certain design affordances, our anthropomorphic mental models could, in fact, diverge from those we hold of other people.

[EXP03]...BUT NOT EXACTLY

This experiment explores the still unsettled matter of equivalence between anthropomorphism and human mental models, with particular focus on whether anthropomorphism can be ‘diverged by design’ from the detected homogeneity. So far in this exploration, experiments have endorsed the conventional tenet of anthropomorphism as a cognitive generalisation from humans to nonhumans, pointing towards a homogeneity of users’ mental models. However, I suspect this may not always be the case and even though certain, if not most, expectations are seamlessly transferred, components such as emotivity or factual knowledge may noticeably differ between the two. Furthermore, this variation has also been hypothesised in *Human Mental Models in Humanoid Robots* (Sau-lai Lee *et al.*, 2005), whereby the authors state that “people might have similar mental models of a person and a robot in one task domain such as mathematical computation, but different mental models of a person and a robot in another task domain”. I am therefore hypothesising that while our estimations of the “kind of person it portrays” heavily rely on our knowledge of other people (demonstrated in the previous experiment), some expectations may vary depending on the task domain and available perceptual information.

Thus, this experiment has been designed to explore two aspects of humanlike AI design: first, whether certain components are indeed transferred non-uniformly (in this case study, I have retained the ‘factual knowledge’ component as a subject of exploration), and secondly, whether a less-literal rendition of the human metaphor resolves some of the issues derived from lumping affordances. This latter scope, with implications on the above-mentioned finding pertaining to stereotype-reinforcement, requires further clarification.

The appropriateness of anthropomorphism in AI design is still subject to ongoing debate, much of which boils down to the unnecessarily uncanny versions of personification the industry is pushing for. Researchers such as Duffy, however, have argued that anthropomorphism is simply a metaphor rather than an explanation and that “the stigma does not stem from the appropriateness in describing the behaviour in terms of anthropomorphic paradigms but rather in cases when such paradigms are used as explanations of its behaviour. Such ‘explanations’ are incorrect, but anthropomorphism is not restricted to only this. It also encompasses facilitation” (Duffy, 2003, p. 181). In his view, ‘anthro-

pomorphism' (note that the term was used to denote not the perceptive act but the design act) should "exploit people's expectations of behaviours rather than necessarily trying to force people to believe that the robot has human reasoning capabilities" (2003, p. 181).

In my view, however, this argument is deficient since all metaphors are explanatory regardless of whether they are explicitly acknowledged or not. A metaphor offers a set of affordances while hiding others by making certain features invisible; they inevitably lead to a suspension of disbelief.⁹ This argument is substantiated by the previous results in which the lumping of information is inherent to the metaphor. Moreover, personified designs are all, ultimately, inherently metaphorical: a machine's 'intelligence' or 'emotions' are fictitious; they are only intended to be *analogous but not identical* to a human counterpart. As such, the question of authenticity in machines is a fundamentally philosophical one. The real challenge, then, is the process of deciding on what constitutes suitable signifiers when it comes to utilising this metaphor and how we draw the lines around misinformation and deceit?

However, while a growing number of publications point towards potential implications on social practices and thus advocate for less-literal executions of the human metaphor in AI, it remains unclear whether a more abstracted rendition would mitigate some of the concerns associated with suspension of disbelief. The following experiment contributes to this debate.

A. METHOD

I. APPARATUS

The scope of this study, intended as a continuation of the previous experiment, is to explore whether varying degrees of humanlike signifiers lead to variation in

⁹ Interfaces, like theatre or storytelling, require users to momentarily suspend critical faculties and accept something fictitious for the sake of entertainment. This is commonly known as the suspension of disbelief and it refers to the willingness of an audience to overlook the limitations of a medium so that it does not interfere with the premise of the story.

users' estimations of an AI's presumed knowledge. Analogous to the previous experiment, the reason for choosing knowledge estimation as an object of analysis was not only motivated by the profound impact it has on our interaction with a system, but also by the idea that estimates of an AI's ability is a concrete indicator of our mental models of it. However, in contrast to the previous experiment, in which measurements were achieved through indirect extrapolations from participants' behaviour, in this case, subjects were overtly required to approximate the device's knowledge of a general domain: the animal kingdom.¹⁰

This investigation was planned to examine the hypotheses that users' expectations and representations of a systems' knowledge would change depending on the type/degree of personification and would differ from their expectations of a human's knowledge of the same domain. In other words, it was expected that the choice of less humanlike signifiers would counteract some of the lumping of affordances noted in the previous experiment.

To test this, participants were asked to play an animal guessing game inspired by Azalea Labs' *Animal Game*¹¹ (a popular Alexa skill in which users are required to keep an animal in mind, which the AI agent would then guess by asking a series of questions), in two different experimental conditions, programmed to depict the two extremes of personified affordances.

In the first experimental condition, the interaction was designed in a literal manner, intended to emulate a human-to-human interaction as closely as possible. In this scenario, while the AI agent was imbued with signifiers of both magnitudes of humanness – Uniquely Human (UH) and Human Nature (HN) dimensions (Haslam,

¹⁰ This choice of domain has been motivated by its universality, in trying to eliminate the potential of it being perceived as specialised knowledge that might lead participants to miscalculate the device's capability in comparison to that of another human being's.

¹¹ The game description says: "Test Alexa's animal knowledge by having her guess an animal of your choice. Alexa will ask you questions to figure out your chosen animal, so make sure you're familiar with the basics! What color is it? Can it fly? Does it live in hot climates? Your answers to these questions will help Alexa narrow down the animal possibilities and make the best guess. Will you be able to stump Alexa, or will she be victorious in the end? Alexa knows over three hundred animals and can also tell you some interesting facts about them. Did you know that rats laugh when they are tickled? And that no polar bear has ever met a penguin?"

Bastian and Bissett, 2004; 2005)¹² – the emphasis was placed on the latter, represented by primary emotions such as sociability and warmth. Therefore, the interaction was designed with the perceptual information of joy, enthusiasm, and surprise when presented with ‘unexpected’ turns of events.

In the second condition, the interaction was designed in a less-literal manner, only broadly resembling aspects of a human-to-human interaction. In this scenario, the AI agent was designed with signifiers from only the Uniquely Human dimension of humanness, represented by aspects such as intelligence and intentionality. For this, the interaction was planned to appear task-driven in a no-nonsense, mechanical manner intended to circumvent the potential affordances of HN, such as empathy, sociability or warmth.

Using this approach, these two AI agents were designed to flag the divergence generated by the type/amount of humanness when constructing new mental models, as well to test a hypothesised difference between people’s anthropomorphic expectations versus their schemas of other people performing the same task.

II. MATERIALS

To ensure that participants were forming new mental models instead of relying on pre-existing ones, this experiment required the design of a new device powered by two distinct AI agents. Given the degree of freedom entailed in the nature of the game, applying a Wizard-of-Oz method would have been impractical, and thus the entire experiment was planned to rely on Amazon’s Echo platform.

¹² “Haslam *et al.*’s research into conceptions of humanness, as manifested in personality traits, suggests that there are two distinct senses of humanness: a species-unique sense, termed *human uniqueness* (HU) and a species-typical sense, termed *human nature* (HN). Human uniqueness, by definition, refers to those traits that distinguish humans from animals, exemplified by refinement, moral sensibility, self-control, and rationality. Uniquely human traits are judged as acquired through learning, as requiring maturity for their expression, not prevalent in the population, and culturally specific. In contrast, HN reflects the biologically-based *human essence* – the fundamental or essential attributes of the human species—some of which may be shared with animals. Traits judged as part of HN reflect emotional responsiveness, prosocial warmth, cognitive openness, and individuality. HN traits tend to be regarded as innate, prevalent within cultures, universal across cultures, and positive or socially desirable.” (Wilson and Haslam, 2013) – see also Appendix E.

However, to increase the probability of generating new mental models, the device was stripped of its staple affordances, firstly by redesigning its outer cover and secondly by programming two separate skills that conveyed very different personas from Amazon's proprietary Alexa. As a result, the digital interaction was purposely designed based on the following criteria:

- + The first persona was designed with the intent of conversing in a highly humanlike manner, resembling some of the most literal renditions of AI personification. As such, in designing the perceptual information that specifies the affordances of uniquely human characteristics, this persona was devised to employ expressions such as: "I'm having so much fun trying to figure this out"; "Wow this is a tough one"; "Amazing choice!".
- + Conversely, the second persona was planned with the intent of 'whitening the black box' (Glanville, 1982), resembling a machinelike personality, designed to specify the affordances of factual competence and predictability. As perceptual information, the 'transparent' interface employed expressions that conveyed the algorithmic steps it took to identify the solution to the quest, such as: "Narrowing down search criteria to exclude mammals"; "Retrieving database of green reptiles".

Correspondingly, the physical form has been deliberately designed based on the following criteria (also see Appendix J):

- + The form had to broadly resemble existing personal assistants to maintain conventions and avoid confusion over identity and function yet be different enough to encourage the formation of new mental models.
- + As one device was employed in both experimental conditions, the form had to accomplish two different functions: for the first experimental condition, it had to appear playful, featuring animal-like cues as a means of visually reiterating the scope of the interaction, whereas, for the second part, it was required to have a degree of accuracy/transparency, by presenting users with a better understanding of the system's components (Figure 19 & Figure 20).
- + In meeting the physical requirements of the first experimental condition, the shape featured animistic perceptual information such as tapered legs and an overall organic shape.

+ In meeting the demands of the second condition, the overall design was similarly dictated by the space and functionality requirements of Amazon's Echo internal modules. As a result, since Amazon's Echo Dot is comprised of three main sections (a small speaker at the bottom, the motherboard in the middle and an LED + sound card at the top), the design was intended to separate these three components into individual layers, as a means of signifying through design, the simplicity of the system, while still maintaining the joyful quality required in the first part of the experiment (Figure 21). The intention here was to utilise a binary of perceptual information conveyed through the design of a single device.

+ Additionally, the button layout was dictated by Echo's soundcard requirements positioned at the top of the device, allowing for the device to still fully function as an Amazon Echo, although stripped of most of its physical affordances.

+ Both experimental conditions were performed with the white version of the device, the black version being a preliminary prototype included here solely for explanatory purposes.

III. PROCEDURE & MEASUREMENTS

To safeguard against inconsistency between experimental conditions, each persona's performance had to be comparable, taking no more than five to six questions to arrive at a solution. To achieve this, participants were required to choose a card from the poll of the multiple choices that have been pre-designed to have similar levels of difficulty. They were then briefed on the apparent purpose of the experiment – *'that of user-testing a novel AI gaming companion'*, after which they were allocated to one of the experimental conditions.

In condition 1, participants were required to select a card and answer the questions posed by the highly humanlike AI agent until it offered the correct solution to the quest. After finalising the game, participants were given a short digital survey, further described below.

In condition 2, participants were required to select a card and answer the questions asked by the robotic AI agent until it guessed the answer to the quest. However, unlike in the previous conditions, after finalising the game, applicants



20



1 LED
+ SOUND CARD
COMPARTMENT

2 MOTHERBOARD
COMPARTMENT

3 SPEAKER
COMPARTMENT



were also presented with a small animation depicting the assembly process of the device's three main hardware components. The purpose of the animation was to offer participants a perceptible sense of control and understanding over its inner workings. After watching the short animation participants were presented with their version of the survey.

The survey (see Appendix H) was comprised, in both cases, of three brief questions: the first two intended to flag possible deviation between users' mental models of the AI agent versus a human performing the same task; the last question was intended to explore whether the less-literal AI rendition leads to less humanlike false affordances. Informed consent has been obtained after providing participants with ample information about the nature of the study and assessing their comprehension in a written consent form prior to initiating the research activity. For safety reasons, data has been anonymised and stored on RCA's Google Drive.

In summary, in the first section, participants were required to assess the likelihood (in percentage) of their allocated AI agent's ability to identify a set of four animals preselected to reflect varying degrees of difficulty (*giant squid* – easy, *Tasmanian devil* – medium, *pangolin* – hard, *fossa* – expert¹³). In the second section, they were similarly asked to estimate the hypothetical probability of another human being's ability to identify the same set of animals – questions intended to flag divergent mental models. Lastly, in the third section, they were required to grade their satisfaction with the agents' performance on a Likert scale from 1 to 10 – intended to isolate whether their expectations have been appropriately met based on the intended perceptual information.

IV. PARTICIPANTS

14 applicants (8 females, 6 males) participated in the first condition of the study, while 15 applicants (7 males, 8 females) participated in the second condition. They were all fluent English speakers and only 2 reported having owned a personal digital assistant (an Echo Dot and Google Home respectively). Irrespective, given the novelty of the interaction, the possibility of participants forming new mental

¹³ These entries were suggested after performing a quick survey in which 20 people were asked to sort ten animals based on their familiarity.

models when interacting with the “gaming companion” was predicted to be relatively high. As they were overtly asked to estimate their expectations of the device’s performance, the survey was anonymised and randomised to safeguard against demand characteristics. Similarly, no further data was collected on participants.

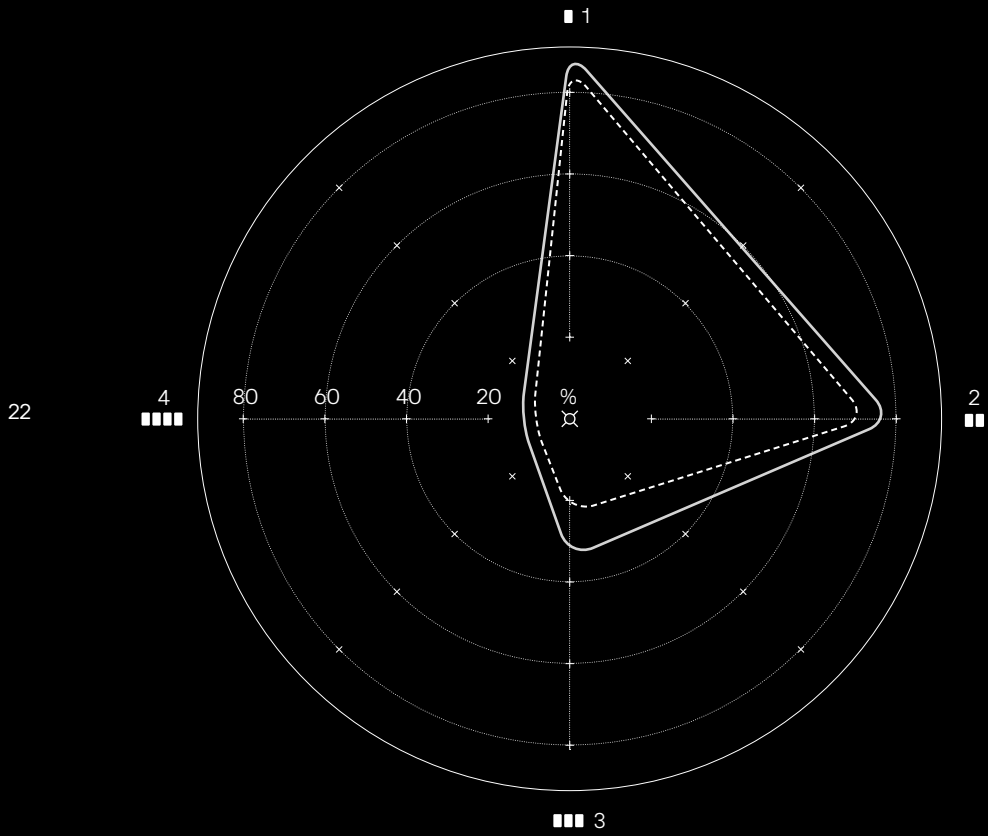
V. DATA ANALYSIS

To recap, participants played an animal guessing game under two different conditions: first with a highly humanlike AI, and secondly a less-literal humanlike AI, after which they were asked to estimate the likelihood of their agent guessing a set of four animals of varying difficulties (detailed results in Appendix I). The same question was then repeated to assess the likelihood of a hypothetical human guessing the same set of animals. In each experimental condition, the score for each animal was averaged across all the participants as follows: first, for their estimation of their allocated AI agents and second for their estimation of a human’s knowledge. These average scores were then rounded to the closest integer and recorded on a radar graph – to visually illustrate overlaps and discrepancies in knowledge estimates (Figure 22 & Figure 23). Lastly, for the final question on participants’ satisfaction with the game companion, the scores were averaged between all participants across the two experimental conditions, acquiring two average satisfaction/likeability scores for each agent.

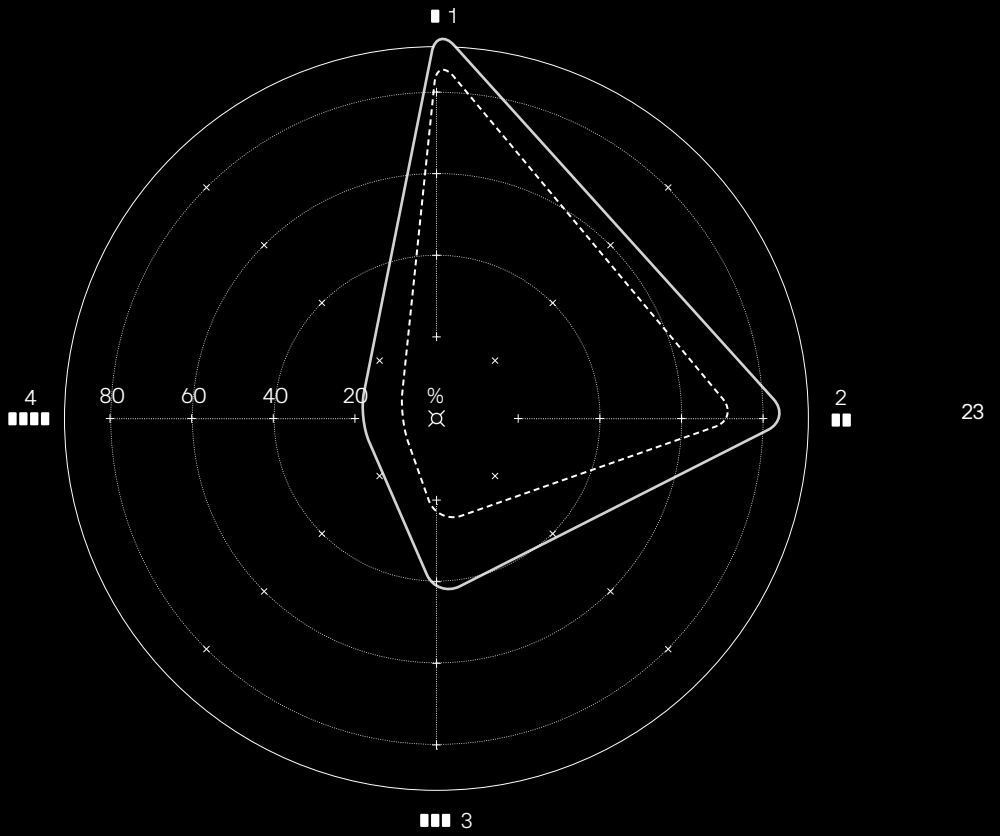
VI. RESULTS

EXPERIMENT 1 (LITERAL CONDITION):

As is visible in the radar graph (Figure 22), participants’ expectation of the AI agent very closely overlaps with the area denoting their expectations of a human performing the same task. This arguably indicates that in this experimental condition, yet again, participants utilised their inferences of other people as an anchor for induction when constructing their expectations of the agent. As a result, the animals that were thought to be recognisable by most people were also deemed as more recognisable by the AI, while their knowledge estimations



- - - - HUMAN KNOWLEDGE ESTIMATION ——— HUMAN KNOWLEDGE ESTIMATION ■ EASY ■ ■ ■ HARD
 ■ ■ MEDIUM ■ ■ ■ ■ EXPERT



----- HUMAN KNOWLEDGE ESTIMATION

———— HUMAN KNOWLEDGE ESTIMATION

■ EASY
■ MEDIUM

■ ■ ■ HARD
■ ■ ■ ■ EXPERT

dropped proportionately for both the human and the AI with increasing difficulty.¹⁴ By looking at the graph, the next distinguishable observation pertains to the area covered by each estimation, with participants' estimation of the agent's knowledge as being slightly larger than that of a human, though, perhaps, not as large as expected.

EXPERIMENT 2 (METAPHORICAL CONDITION):

Echoing the findings from the previous condition, the two areas denoting human and the AI agent's knowledge estimation point towards the now predictable similarity in user's construction of mental models (Figure 23). This is evidenced by participants' analogous process of inverse correlation between knowledge estimation and the perceived difficulty in identifying the animal. However, in this condition, there is notably a more evident discrepancy in the overall area covered by each graph – with AI's knowledge estimation extending slightly beyond the area corresponding to human knowledge estimation.

Hinting at the hypothesised divergence between anthropomorphic perception and its corresponding human counterpart, the gap between these two knowledge estimations possibly alludes to participants' overall higher trust in the AI's ability to predict the entries. Similarly, when overlapping the area covered by the metaphorical and the literal AIs, we notice a very similar, though slightly narrower, gap in knowledge estimation (Figure 24). These aspects will be further examined in the 'Discussion' section. Finally, averaging the scores obtained in the last section on users' satisfaction with the interaction leads to an overall score of 9.2 for the literal condition, and 6.9 for the metaphorical one – results which pose some challenges to the relationship between transparency and user satisfaction (see Appendix I).

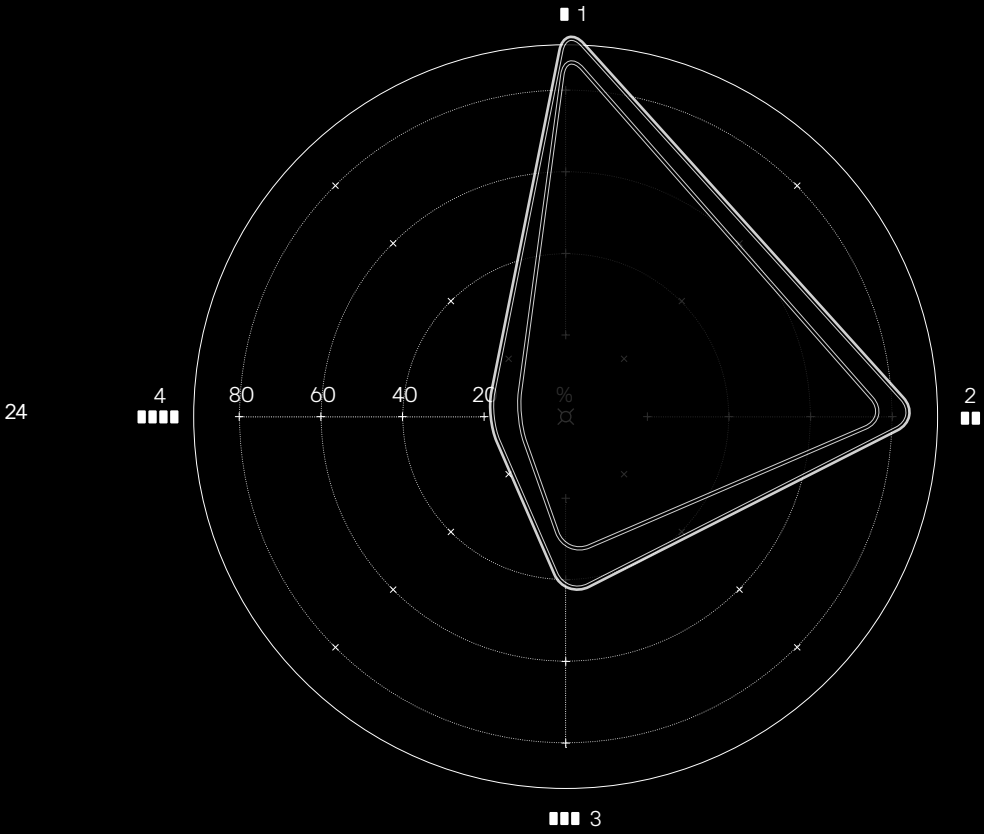
¹⁴ While we notice a similar estimation between human vs AI knowledge for the easy, medium and expert entries, there is a slight discrepancy when it comes to the "hard to recognise" animal. This discrepancy might be due to a different expectation of the AI when it comes to factual knowledge – most evident when the animal to be guessed is lesser-known but not entirely obscure (in the obscure/expert entry, the score becomes, yet again, equivalent to that of another human's).

B. FINDINGS

Adding to the findings from the *Just Like Me* investigation, these results compound support for the hypothesis that regardless of the type/degree of personification (and therefore extent of perceptual information), users' anthropomorphic mental models (and thus expectations of affordances) are extrapolated from their knowledge of other people. Substantiating the main premise, anthropomorphic perceptions seem, indeed, to build on users' own mental states as a basis for induction – interpretation which lends to a discussion of whether anthropomorphism can, in fact, be effectively 'diverged by design' from its detected homogeneity with human mental models.

As such, these results may be interpreted in two ways: either users projected their own interpretations of the difficulty of a task onto their expectations of the agents' knowledge, or their trust in their own ability to play the same game decreased with perceived difficulty. If the latter is correct, we could be tempted to interpret the results as not having lent any insights into users' mental models, but we may be overlooking one important aspect: that irrespective of whether these results were a representation of participants' own ability to answer the questions, they still estimated the metaphorical AI's knowledge to be greater than that of the literal AI agent (Figure 24). When compared to the expectations of the literal or to that of another human's knowledge estimations, this ultimately denotes an overall higher confidence in participants' estimates of the metaphorical agent's ability to reliably perform this task.

As we have seen in the literal condition, in which the agent was modelled to emulate a human-to-human interaction as closely as possible, participants' expectations of it appear to be equivalent to their expectations of another human being when it comes to simple tasks, and only slightly higher when it comes to difficult (but not expert) tasks. This variation might be due to users having a higher estimation of a computer's knowledge of factual information and therefore trusting marginally more the agent's ability to identify difficult entries. However, this interpretation cannot be established with certainty. Given the small sample of participants, this slight inconsistency is not nearly large enough to safely draw this conclusion and for the most part, we can assume that, in the literal condition, users' expectations of the AI are similar to their expectations of other people.



Conversely, this was not the case in the metaphorical condition, in which the AI interaction was designed to communicate its 'algorithmic processes' with intelligence and intentionality while being stripped of the perceptual information of primary emotion or warmth. In this scenario, users visibly favoured the AI's odds of recognising an entry when compared to those of another human, despite also being introduced to the device's uncomplicated inner workings.

One potential explanation for this might derive from people's socially constructed discourse on machinelike AI agents with vast 'factual knowledge'. In this vein, the perceptual information of an apparent emotionless persona is possibly directly correlated with higher knowledge estimates. This rationale goes hand in hand with the opposing effect in which the perceptual information of emotivity and warmth may result in reduced knowledge approximations, perhaps through association with humans' limitations when performing the same task.

Another possible explanation might pertain to the conceptualisation of the AI agents. In 'whitening the black box', participants were presented with information on the agent's rationale for identifying entries and could have perhaps reasoned that the same processes could be successfully applied to more complex tasks. Correspondingly, in the literal condition, without means of extrapolating how the agent was executing the task, their trust in its performance dropped as the difficulty increased. As such, despite being introduced to the simplicity of its hardware, gaining insights into the agent's data processing could have led to an impression of control and confidence with its performance. If this is indeed the case it could mean that more 'transparent' design conceptualisations of AI would lead, in factual knowledge scenarios, to an increase in users' expectations of a personified system.

One could be tempted to advocate for the established relationship between transparency and trust in AI design as a panacea for its known shortcomings, but in doing so, may be overlooking one noteworthy aspect. While participants indeed favoured the metaphorical conceptualisation in terms of placing their trust in its performance, they scored considerably higher their satisfaction with the literal interaction. In short, participants did not necessarily enjoy the interaction they trusted the most.

To interpret these findings, we could, once more, employ the lens of cognitive dissonance. As such, in the gaming context of this interaction, enjoyment is

arguably directly correlated with an AI's ability to baffle and astound; as a result, transparency might not be a desirable or even expected affordance (at least not subconsciously).

Moreover, social constructions of what a game entails bring about expectations of emotion, enthusiasm and warmth – attitudes that have been purposefully left out from the metaphorical design, consequently challenging the accepted social norms related to the context of the interaction. Trust and likeability, then, do not automatically go hand in hand, and as users' expectations are pushed and pulled by varying social factors, simply taming humanlike design conceptualisations may not necessarily lead to frictionless interactions.

DISCUSSION

A NEW ROLE

The observed discrepancy between anthropomorphism and its human-world equivalent when it comes to less-personified designs poses the question of whether anthropomorphism is indeed anchored in our human mental models of other people. As seen, one could be tempted to contend that people incorporate other mental models, such as those of machine automation, whereby their expectations of a computer's implementation of a computational task is often believed to outperform its human counterpart. However, if people had indeed employed different schemas, we should have expected to find consistency throughout their estimates of the agents' likelihood of identifying entries as opposed to the observed variation between familiar versus unfamiliar animals. This variation testifies to users' mechanism of constructing mental models still arguably relying on the same cognitive tools utilised in predicting other people's behaviour. But this poses a further question: given the perceived homogeneity, what is the justification for the observed discrepancy?

One possible explanation might be with respect to a potentially new socially constructed role assigned to humanlike AI agents. This could imply that, albeit still grounded in anthropomorphic mental models, expectations (and therefore affordances) shift in relationship to the performed task. The human equivalent of this relates to the construction of new specialised roles, whereby our role

schemas shift to accommodate an entirely new set of expectations of the humans performing them. A very similar phenomenon might be taking place here, whereby participants may have been relying on altered role schemas, potentially influenced by cultural representations of intelligent machines with vast factual knowledge but limited capacity to display emotions. In this vein, the more robotic AI may have primed new role schemas associated with different expectations of the agent's abilities.

However, as this work is at an early stage further research is needed to catalogue and understand the impact of these potentially new roles. Nevertheless, within the scope of this investigation, it can only be concluded that the acts of designing the perceptual information that specifies anthropomorphic affordances appears to be additionally complicated by the very existence of fluctuating new role schemas with potentially novel expectations unparalleled in our human social reality.

[NON]METAPHORICAL ≠ BETTER

Secondly, this experiment contributes to the debate on whether simply approaching AI humanlike design from a less literal emulation of human-to-human interaction would solve some of the primary concerns associated with anthropomorphism and sequential affordances.

It is my contention that this is not necessarily the case. While we cannot directly draw any conclusions on how less-literal renditions of anthropomorphism impact social practices (such as whether users still resort to stereotypical beliefs), we can note that, depending on the context, users could end up favouring more humanlike conceptualisation. These results are by no means universal, but they do indicate that simply diluting humanlike design cues might resolve certain expectations (such as trust), while equally generating other false affordances in the process. As a result, the use of humanlike design cues is not a fault in itself, but rather a fault of AI designers' often indiscriminate use of *anthropomorphic features* without fully appreciating their implications on perception.

These findings may therefore suggest that in the process of designing humanlike AI agents, there may be no straightforward recipe and simply deciding between entirely humanlike or not at all proves to be challenging. This is an inherent consequence of how our expectations of personified agents are built upon our

mental models of other people, and thus, irrespective of the number of humanlike traits it incorporates, users' anthropomorphic schemas are still very complex and exceptionally volatile.

In concluding this chapter, we are faced, again, with the same limiting assumptions engrained in the choice of methods used to identifying users' mental models of AI agents, since what is expected and appropriate when designing humanlike interactions fluctuates with the plethora of internal and external cues primed by perceived affordances. As emerging from these last three experimentations, the ability to navigate the relationship between personification (concept) and anthropomorphism (perception) is furthermore obfuscated by the perceptual fluctuations stemming from the overabundance of the perceptual information, signified sometimes inadvertently through the design of humanlike agents.

These lumping and fluctuations of perceived affordances are, most likely, a consequence of the socially dynamic system of inferences between the Human and the AI – meaning that for *any* output of a practice governed by a personified frame the resultant object inherits the limiting assumptions of that frame.

As seen from the last two investigations, though the form of personification impacted anthropomorphic inferences it did not manifest in predictable ways. This is perhaps testimony to the sophistication of the socially constructed anthropomorphic mental models that inherently rely on the continuously constructed social reality. Consequently, merely controlling “the illusion of intelligence” may not bypass the complications engrained in interaction between personification (and the choice of signifiers) and anthropomorphic inferences.

On the contrary, in any design frame which foresees a consequential relation between the two, the resultant artefact is intended for *imagined* rather than *real, meaning-constructing* humans who shift their expectations, not only in response to stable internal motivators but also in response to the relational facet of the Human-AI mutuality. However, in exploring these fluctuations, it becomes apparent that a more tailored methodology fit for the investigation of a socially dynamic system is needed. This is discussed in the next chapter.

03

THE HUMAN-AI

Thus far it was noted how the prevailing focus in the AI research community – that of embracing a pragmatist philosophy that prioritises the significance of the resultant action over the construction of meaning – performs poorly when applied to non-obvious or even paradoxical aspects of Human-AI behaviour. The earlier two chapters, similarly, grappled with the view that either humans are the locus of values and truths or that behaviours can be easily codified and extracted from deliberate (or non-deliberate) affordances, external to the social context. Still, in doing so, we have seen how this view is an oversimplification of how the social mechanisms between the Human and AI unfolds – resulting not from a false assumption regarding how our social world works but rather from the conviction that there may not be a social world unfolding, to begin with.

As such, in the chapters on the Human and the AI, I have gradually begun challenging this practice by demonstrating that human behaviour cannot be unproblematically predicted by perceptual theories of anthropomorphism and neither can it be easily manipulated by means of design. The relationship between the two is profoundly contextual and, if it is even possible at all to universalise, cannot be done so effortlessly.

In this vein, work by a number of researchers in areas intersecting the themes of HCI and design research tackles the use of AI in domestic settings, such as: Alex Taylor’s study on machine intelligence (2009); Chalmers & Galani’s emphasis on the temporal and special facets of a context of interaction (2004); or Pierce and DiSalvo’s work on design metaphors (2018), present potential avenues of research for tackling the profoundly contextual nature of the Human-AI interaction.

Nevertheless, following a review of this work and its importance in the realm of research in which I am operating, it is clear that studies such as these, still broadly tackle the Human and the AI as an out-of-sync construct. For this reason, these avenues of research are out of this study’s scope of tackling what is now the last pressing question: *How are expectations continuously constituted through the interaction between the two in practice?*

As such, to circumvent the inherited limitations of an out-of-sync interpretative process, the previous three experiments required a reconfiguration at the level of research strategy – embracing a symbolic interactionist approach that depicts humans as active agents who interact with other AI agents based on their deeply contextual, interpreted meanings. In doing so, I have resorted to what is perhaps

one of the most influential and widely studied phenomena in the history of social psychology and applied it to the (until-now debated) social context of Human-AI interaction – cognitive dissonance theory (CDT).

Thus, in this chapter, I start by presenting a brief description of the theory of cognitive dissonance as it first appeared in Festinger’s book *Theory of Cognitive Dissonance*, followed by delineating the theoretical entities that have contributed to its success in explaining the non-obvious aspects of the interaction. However, in reviewing the evidence bearing to this theory, I shall also discuss its current methodological limitations and why this research proposes a revision of its methodology as necessary to develop the theory further. As such, the methodology will move away from the traditional approach of studying dissonance, towards the symbolic interactionist perspective of seeing meaning to be “mediated through language, defined through the interaction and materialised into action” (Kenny and Fourie, 2015).

In summary, building on the previous experiments in which meaning has been mediated by language, I propose a methodological approach that weaves together the Human with the technological and social elements of AI design, by studying how these elements visually materialise into a landscape of interaction.

HUMAN-AI <> COGNITIVE DISSONANCE

The proliferation of a theory published more than 60 years ago can be traced back not only to the success with which it challenged the long-standing dominance of reinforcement theory (based on reward-based explanations) but also to the very simplicity of its theoretical model. The core notion of the theory is deceptively simple: psychologically inconsistent cognitions produce a negative drive state (dissonance) that motivates individuals to strive towards consistency (consonance). While CDT has been primarily applied to socio-psychological contexts, its implications have been investigated across a wide range of fields (Aronson, 1969).

For two decades after its initial formulation by Festinger (1957), the theory gave birth to a new era of social psychology that combined motivation and cognition in stark contrast to the previously prevailing perspective of humans as reinforcement machines. As such, one of CDT’s greatest strengths was that it frequently

challenged the predictions made by behavioural approaches (Brehm and Cohen, 1962) – with the noteworthy contradiction in demonstrating that individuals also like things for which they suffer (as opposed to only things for which they are being rewarded).

However, despite its vast contribution to research, by the mid-1970s, the allure of the theory began to wane since studying dissonance implied the use of a methodological approach that understandably received a great deal of criticism (Aronson, 1992). Namely, participants were subjected to ‘live experiments’ in which dissonance was induced in real-time, under procedures that usually caused discomfort or sometimes harm to individuals (Metin and Camgoz, 2011). As a result, research moved away towards more straightforward, less time-consuming methodological approaches that were reasonably deemed more ethical, opening a new era in favour of purely cognitive studies (the origins of which can nevertheless be traced back to CDT).

Nonetheless, despite the waning of interest in dissonance research, its formulation remains exceedingly accurate, and many of the theories on cognitive dissonance utilised today are not only derived from CDT but could not have been derived from any other theory, to begin with (Cooper, 2007). As such, a more detailed discussion of any of the more recent theories of self-concept (Steele, 1988; Higgins, 1989) or self-motivation (Kunda, 1990) is beyond the purpose of this investigation, even though they can be integrated within the framework of CDT, despite its relatively dated formulation.

While there are certainly other operational modes for pursuing this type of investigation, for the past three experiments, the CDT framework has proved to be especially suitable for situations in which either users’ mental models are convoluted (most, if not all, anthropomorphic models) or the conceptual models are obfuscated by technological black-boxes (most, if not all, social AI agents).

By studying the interaction between the two through a lens of dissonance and reaction to dissonance I propose a simple account for observed phenomena implying no more than a small number of postulated theoretical entities. These have been refashioned from the original CDT to lend to a more flexible and intuitive methodology with emphasis not on motivation and cognition but rather on the meaning-making process, mediated by language.

However, before pursuing a CDT-derived framework of analysis, I must first address the obvious methodological differences between dissonance research and this research. Namely, in this research I do not adhere strictly to the methodological tactics of CDT, instead I am relying on data and observations of behaviours naturally occurring in participants' interactions with AI. In other words, dissonance was not the object of the study but rather a by-product of the interaction that materialised after interpreting the behaviours on display. Therefore, while the goal of this study is not to refine or expand the theory of cognitive dissonance, CDT provides an invaluable foundation for building a new predictive model of interaction between the Human and the AI, one that requires a methodological reformation.

A. THEORETICAL ENTITIES

I. INCONSISTENCY

An assumption of inconsistency between the Human and AI's social interaction is core to the questions and hypotheses guiding this research, and such inconsistency has indeed been observed and outlined in previous experiments. Now, in discussing CDT, we have the vocabulary with which to more explicitly define it. Festinger characterises inconsistency as the psychological discomfort arising when an individual holds two or more inconsistent elements of cognition.¹ It should be noted that, although Festinger's account of cognitive elements is justifiably ambiguous since his work was conducted at the "dawn of the cognitive revolution in psychology" (Kaaronen, 2018), it still overlaps with what constitutes a mental model throughout this investigation. To justify this overlap of terms, in the author's own account, these cognitive elements form a "mirror, or map, of reality", a reality that "may be physical or social or psychological, but in any case, the cognition more or less maps it" (Festinger, 1957, p. 10).

¹ Festinger defined "elements of cognition" to represent what "a person knows about himself, about his behaviour and about his surroundings", in other words, they represent not only elements of self-knowledge such as attitudes, needs or feelings but also causal and environmental conditions referring to "the world in which one lives" or "what is where, what leads to what" (Festinger, 1957, p. 9).

II. TENSION

As a second theoretical entity in CDT, tension is defined as the state resulting from an imbalance, therefore, a natural by-product of inconsistency. In dissonance theories, tension is understood to be the negative drive state that leads to the need for achieving consonance.

Tension, however, may arise from any permutation of inconsistency such as: a mismatch between two related attitudes, a mismatch between belief and behaviour, as well as a mismatch between a cognitive element and observed reality. This mismatch between one element of cognition and observed reality forms the basis for the previously defined *false affordances*, which, by definition, in this framework naturally result in inconsistency and, therefore, tension.

In CDT, the magnitude of dissonance depends on the importance the individuals place on the cognitive elements: the greater the importance of these elements, the greater the magnitude of tension once they become dissonant. Thus, according to Festinger, we should be “able to observe some manifestations of these pressures” (1957, p. 11), denoting that tension is at the heart of “enabling measurements of dissonance”.

III. STRATEGY

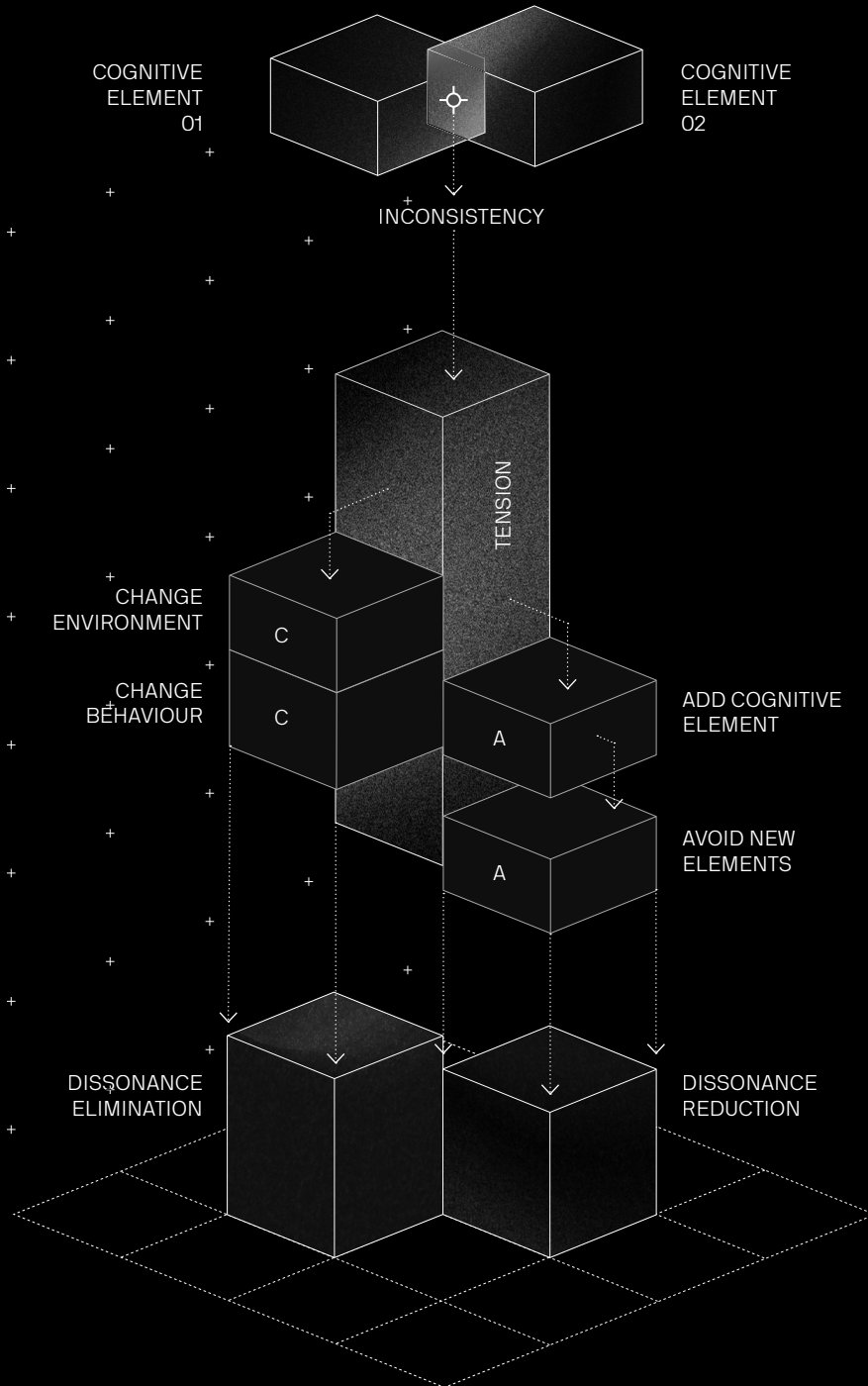
Finally, the central aspect of dissonance theory is that humans have a natural tendency to manage these cognitive elements in response to reality in a way that alleviates inconsistencies. In other words, one of the main tenets of CDT is that “the reality which impinges on a person will exert pressures in the direction of bringing the appropriate cognitive elements into correspondence with reality” (Festinger, 1957, p. 11).

Therefore, *inconsistency* is a motivating factor in itself, exercising a *tension* that is psychologically uncomfortable and will consequently prompt the individual to achieve consonance through one of the following three strategies: *elimination*, *reduction* or *avoidance*. These strategies can be further broken down into attempts to change either behavioural or environmental cognitive elements, adding new cognitive elements to reduce inconsistency, or avoiding new elements altogether (Figure 25). To give a now canonical example, a smoker who

has just learnt about the effects of smoking may achieve consonance through one of the following strategies: changing their behaviour by quitting smoking (eliminating dissonance); adding new cognitive elements such as “you can die from anything” (dissonance reduction); or changing their knowledge about the effects of smoking, e.g., “smoking is not bad” (avoidance).

In this vein, we could now re-evaluate insights derived from the previous three experiments through the lens of the proposed theoretical framework. Then, when overlapping these investigations with the diagrammatic mechanism suggested above, we may be able to extrapolate the following:

1. The first experiment revealed people’s tendencies to amend either their personal evaluation or their evaluations of the source (AI agents) in the direction of least inconsistency. As a result, it was noted how a positive evaluation of experiences with their personal assistant (higher satisfaction/ understanding), was associated with a similarly matching opinion of the agent – where a positive attitude was directly correlated with the tendency of ascribing the AI more humanlike affordances (changing a behavioural cognitive element; elimination). The opposite was also true: when dissatisfied with its performance, these affordances did not exist to begin with.
2. The second experiment exposed another strategy of managing possible inconsistencies between people’s belief of an AI agent’s knowledge and their behaviour towards it. As a result, this investigation revealed a direct association between people’s belief and hence expectation (that an individual should have more extensive knowledge of local landmarks) and the behaviour employed in response (i.e., the amount of time spent describing each landmark). This is in line with adding the new cognitive element of AI’s presumed knowledge to achieve dissonance reduction.
3. Finally, the same framework may be employed to clarify participants’ evaluation of the literal agent as more engaging, despite its noticeable lack of transparency. This was arguably the result of matching their expectations of the context of the interaction (which sees gaming as being a form of entertainment) with their expectations of the agent’s behaviour. As a result, the value of “enjoyment” was constructed contextually, and the apparent affordances changed in response to it (changing a behavioural cognitive element; elimination).



As such, we may notice how the deeply social expectations of the Human-AI interaction do not remain the same when moving from one context to the next. However, while understanding these histories and dynamics is central to this investigation, isolating what ends up being an inconsistency in the gap between the Human and the AI may not be such an easy feat. Based, thus far, on previous attempts to identify the potential false affordances of the Human-AI interaction, I have found that delineating what may constitute a deviation or inconsistency prior to the interaction turned out to be a cognitively daunting task. This may be because, as indicated experimentally, one personification signifier can lead to a series of other unexpected elements of cognition, which can hardly be anticipated in a practice governed by an anthropomorphic frame.

Because of the inherent limitation of the strategy employed thus far, I started interrogating the rigorousness with which I was employing this cognitive exercise – that nevertheless involved the analysis of only but a few affordances, as opposed to the evaluation of an entire socio-technological experience. This motivated the search for finding alternative strategies of revealing the hidden structures of the interaction.

As a result, drawing from both my background in interaction design and data visualisation and the requirements of condensing what seemed as a rightfully complex phenomenon, in the ensuing section I explore the possibility of visualising tension and inconsistency as being locked away in the more subtle behavioural cues that arise moment by moment in the interaction. To do so, I am proposing a visual language – an interactive topography that could aid the process of delineating the behavioural outliers of this socio-technological setting with the purpose of making *inconsistencies* more visible for interpretation.

[EXP04] INTERSPACE

This experiment explores a digital method of analysing human behaviour based on observations of everyday conversational practices with AI agents (in this case, by looking at linguistic productions as the primary vehicle of interaction with humanlike technology²).

At the heart of any social interaction, however, lies an enormously complex act of transferring meaningful information about who we are and what we think, feel or sense via an array of verbal and non-verbal channels. Human dialogue is sufficiently multifaceted that identifying conversational patterns among the ambiguities of natural language, speech characteristics and paralanguage may well be impossible. For this very reason, communication research is an enormous undertaking drawing from many disciplines, each placing their unique weight on either the verbal, with a focus on language, syntax (Chomsky, 2002) and social markers in speech (Ryan and Cacioppo, 1981); or non-verbal, with a focus on speech styles, paralanguage and overall body-language (Patterson, 1983). This fragmented approach stems from the impracticality of scrutinising the whole interaction at once.³

Much can be said about each of these dimensions independently, as they distinctively lend different perspectives on interpreting human behaviour. In this case, however, identifying inconsistencies appears to require a broader-picture tactic. Accordingly, in this experiment I start from the idea that text and non-text

2 Certainly, other sources of data could have been used instead, such as non-verbal behavioural maps or cognitive mapping. However, given the heavy emphasis this research has thus far placed on communication and language, when paired with the necessity of studying these interactions as naturally occurring outside of experimental conditions linguistic production seemed the richest source of information within the context of this investigation.

3 In this vein, the social psychology of language tends to overlook the text of a communicative act, instead favouring analyses of speech style and non-verbal communication. It further tends to break down exchanges into component parts and then reconstruct more complex events from the interaction of different channels (Hogg and Vaughan, 2017). Similarly, ethnomethodological methods such as 'conversation analysis' lean towards transcriptions, logically favouring text over non-text as a means of studying oral language practices.

dimensions are inextricably linked, together conveying meaning and reflecting attitudes, motivations and intentions – elements that often only emerge when considered at a discourse level.

As a result, the ensuing experimental method considers the entire conversation as a unit of analysis, looking not only at what is said, by whom it is said and for what purpose, but also in what way the meaning of the text is influenced by speech styles and context.⁴ This is because it was not particularly expected to find pattern deviations at a granular level by isolating verbal or non-verbal cues, but rather by studying the interaction through a fundamentally interactionist approach (focused on the construction of meaning).

Although it can be argued that the same could be achieved by studying detailed transcripts of an interaction (in a Conversational Analysis approach), an earlier attempt pursuing this approach revealed that this style of analysis poses some serious limitations (see Appendix K). In particular, the more we strive for accuracy and detail the further it decreases the method's readability and, thus, the quality of the evaluation is at the mercy of a compromise between usability and exhaustiveness.⁵ The visual representation is, consequently, a result of this need to bypass the challenges of transcription. Additionally, through this visual coding, the task of inspecting the conversation at once can be made more intuitive by eliminating the necessity of switching back and forth between transcript details.

The remainder of this chapter will focus on how interactions between users and AI can be studied as abstracted representations of patterns of conversations. Nevertheless, achieving this posed its own challenges such as: deciding on what data is prioritised from the plethora of conversational dimensions, how is this data collected, and what the criteria are for designing a distinctive visual language when establishing the 'meaningfulness' of this approach. In the following pages, I will provide further clarity with regards to the encoding strategy.

4 "The idea of looking at the entire conversation as a unit of analysis has parallels in the world of machine learning, where natural language processing has shifted from looking at very localised relationships between words to looking at the shape and significance of elements of a whole sentence. See: (Wolf *et al.*, 2020)(used most visibly in GPT/3)". Cited from an email exchange with my supervisor, Richard Banks.

5 This is because translating behaviour from transcripts implies having to follow an interaction in a similar manner to the one in which it took place, which becomes a virtually impossible cognitive task with a transcription full of behavioural annotations.

B. EXISTING WORK

A review of existing literature has revealed several antecedents for the graphical depiction of conversations, some serving as inspiration moving forward. In considering specifically those that have influenced and motivated the approach used here, I will first mention the studies that could be described as visualising content alone. In this category we find studies such as Havre, Hetzler and Nowell's *ThemeRiver* (2000), a digital tool that visualises data-based word usage frequency. *ThemeRiver* depicts variations horizontally, laid out with varying thicknesses according to word occurrences, giving the impression of a flowing river – an idea that inspired the early attempt to visualise conversations as “flowing” topologies (see Appendix M). Similarly, Wise *et al.* (1995) visualise text content from various sources to create spatial representations in the shape of geographic terrain (literally, a landscape of themes). They claim that “the result is an interaction with text that more nearly resembles perception and action with the natural world than with the abstractions of written language”. This idea of a landscape, that more readily taps into our perceptual intuition, inspired the choice of an interface resembling a physical space. Further, Rohrer *et al.* (1998) propose a method of exploring written non-visual information, by specifically mapping text onto three-dimensional blobs that morph and form clusters based on “inter-document relationships”. While all these conceptual categories inspired the proposed approach, what separates them is the preoccupation with the text dimension alone, distinctly pursuing a different set of research questions.

The aim of using graphics to convey patterns of activity and dynamics of conversation has been, however, successfully tackled in two of MIT's online chat rooms: *Chat Circles* and *Bubba Talk*. *Chat Circles*, unlike other graphical chat rooms that adhere to realism, uses abstract shapes to depict identity and patterns in conversation. In the authors' words: “each participant is represented by a coloured circle on the screen in which his or her words appear. The circles grow and brighten with each message, and they fade and diminish in periods of silence, though they do not disappear completely so long as the participant is connected to the chat” (Viégas, 1999). Similarly, *Bubba Talk* is a visualisation technique used to represent various aspects of an interaction and to give an overall “impression of the spirit and timbre of the conversation” (Donath, Karahalios and Viégas, 1999). Significantly, with these two examples, it is not the avatar-based graphical chat system that is remarkable, but rather the visual emergence of a social structure, which served as inspiration moving forward.

Similarly, a wealth of visual representations of the structure of text can be found outside the realm of academic research. When discussing visual encoding of text, visual representations such as Posavec & McInerney's *The Evolution of the Origin of Species* (2009) or Harrison and Romhild's *Visualizing the Bible* (2008) are further examples of making visible something that is locked away in the words of a text.

These examples inspired some of the coding procedures used to analyse data in this study. However, to my best knowledge, at the time of writing, research on both surface-mapping based on linguistic productions (regardless of the nature of the agents) and visually mapping behavioural dissonances are either lacking or non-existent. Additionally, despite the growing incidence of language-mediated human-computer interaction, there is a considerable lack of methods catering to the underlying social patterns of this emerging socio-technological landscape.

C. METHOD

The suggested method, InterSpace (short for Interaction Space), uses the quantifiable elements of speech to generate flexible, procedural shapes or blobs. This generation of blobs needs no more than a small set of parameters to dynamically spawn and encode the rich details of an interaction. Since visualising the interaction requires mapping the abstract onto the physical, three-dimensional shapes provide both the required level of abstraction as well as the level of complexity sufficient to dynamically encode a progressive set of non-spatial elements (such as paralanguage, turn design or social structure on different layers of interaction).⁶

Following existing research on language, this method first breaks down Human-AI interactions into components, starting from one turn as a unit of analysis and then gradually constructing an artificial landscape that encrypts further layers of social complexity. In the process of selecting which data to visualise, multiple criteria were considered, but the outcome was more so dictated by the aim of the project. As such, focusing on patterns of response and participation, fluctuating moods and overall synchronisation proved the more revealing method of materialising the ebbs and flows of a conversation. These were also hypothesised to

⁶ Naturally, prior to arriving at this method, multiple other visualisation attempts have been trialled – which, for clarity of structure, have been included in Appendix M.

highlight the potential incongruities or dissonances that arise on the background of an emerging social structure.

The choice of what data to be visualised, in turn, influenced its method of collection. With data needing to be as naturally occurring as possible, (devoid of the many factors that may influence social interaction, including the presence of other people), this method resorted to recordings submitted by familiar participants that have agreed to engage in an interaction with their personal digital assistant from the comfort of their homes.⁷ To avoid further undesired manipulation, they were given the very open-ended brief of achieving a simple task with their AI assistants. They were then asked to upload to RCA's cloud storage their first audio recording (or all the subsequent recordings), regardless of whether the task had been successfully achieved or not.

These audio recordings were then passed through PRAAT (*Praat*, 2022), a speech analysis software used by linguists to label, segment, annotate and generate transcriptions of speech recordings. PRAAT was further used to break down the content of each recording into several acoustic analyses, including spectrograms (visual representations of the high and low frequencies); pitch contours and formant contours (Boersma and van Heuven, 2001).

These spectrograms, which functioned as a first step towards using visual representations for pattern recognition, rapidly became virtually impossible to navigate at the level of discourse. This insight gives further weight to the requirement of automating the sense-making process of flagging patterns and inconsistencies based on linguistic data.

Next, in the process of evaluating the method's efficiency, I further encountered the challenge of having to continuously evaluate both the choice of data as well as the visualisation language – either of which could be freely altered, independently of the other. To account for this, determining which approach was best suited to be carried on, came down to meeting the demands of meaningfulness versus readability.

7 The choice of using familiar participants for this project can be substantiated by the need to reduce demand characteristics. In this peculiar case, familiarity appeared to be a great contributor to less of a need to “succeed”. For this very reason, participants were also asked to record their conversation with their personal assistant – creating a context as naturally occurring as possible

Finally, when tackling these issues, the premise that a conversation is not merely an exchange of data, but a complex social landscape (in which the words people say are only part of the overall message), has been the key factor in deciding on the visualisation approach. As such, Human-AI conversations are broken down into three different layers of abstraction, derived from existing research on language (Hogg and Vaughan, 2017), each entailing a different choice of representative parameters and visualisation methods. Together they construct an interactive topography that provides a subtler and more nuanced overview of the interaction:

- + Language – interpreting data pertaining to utterances, duration, gaps, lapses or overlaps – underlining personal as well as overall balance in sequence design and structure.
- + Paralanguage – interpreting data such as volume, pitch or voice breaks – indicating mood shifts and changing attitudes.
- + Structure – interpreting the in-between, the silences, the combination of parameters, reflecting the social landscape and highlighting potential deviations from average or points of tension.

These three layers required their own distinct sets of representative parameters and visualisation techniques that nevertheless had to work in tandem to deliver a rich yet coherent interface for exploration.

I. LANGUAGE

In generating each landscape, I defined a simple method that takes the leitmotif of any conversation analysis – turn design – and maps it onto the simplest flexible, procedural shape, considered here as a ‘unit of speech’. Playing on the historical representation of turns in comic strips or graphic novels as dialogue or speech balloons, the method maps each turn onto a three-dimensional blob in space, based on a small set of parameters further detailed below. Each speaker is assigned an individual colour throughout the landscape as a means of distinguishing their overall contribution to the discourse and sequence design. In this sense, for the generated case studies used in this analysis, the user is rendered in grey throughout the conversation while the AI agent is rendered in black.

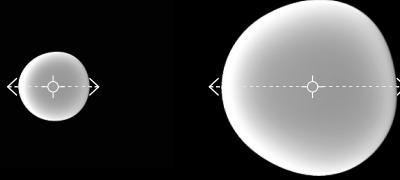
Among the four parameters used to control the generation of shape, duration of speech is perhaps the most noticeable. Echoing a conventional expectation for bubbles to increase with contribution (or with word count in the case of comic strips), in this method, bubbles grow based on the overall time spent on completing the turn. This is reflected in the diameter of each bubble, which is equal to the number of seconds it took before a pause intended to select another speaker (Figure 26, 1). The choice to use seconds rather than word count was motivated in large part by the consideration of how much ‘space’ a speaker occupied in a conversation. Indeed, while this method would give equal consideration to a turn uttered quickly and a slowly delivered turn transmitting less content, the space occupied by these turns was ultimately of more value to the analysis. It should also be noted, however, that the decision was also influenced by seconds being more accessible to the utilised algorithm than was word count (see Appendix O).

Next, for an interaction to make sense, it must be read in a certain order. Unlike in comic strip and graphic novel conventions whereby readers should follow the convention of reading the bubbles stacked in the order of the language (so from left to right and up to down), in this method bubbles appear on a timeline of a conversation, with empty space occupying the expanse of silence in between turns. Thus, the interaction is still read in the order of the language, from left to right, but solely on a horizontal X-axis: the location in space and the size of each bubble is directly related to when and for how long a speaker entered the conversation, respectively, with bubbles ending their growth the moment the speaker stops talking. This allows for empty space to ‘form’ instead. While the generation and placement of shapes are determined by turn duration and timing, the generation of detail in each shape is subject to three other paralanguage parameters: pitch, voice breaks and volume. These are further detailed below.

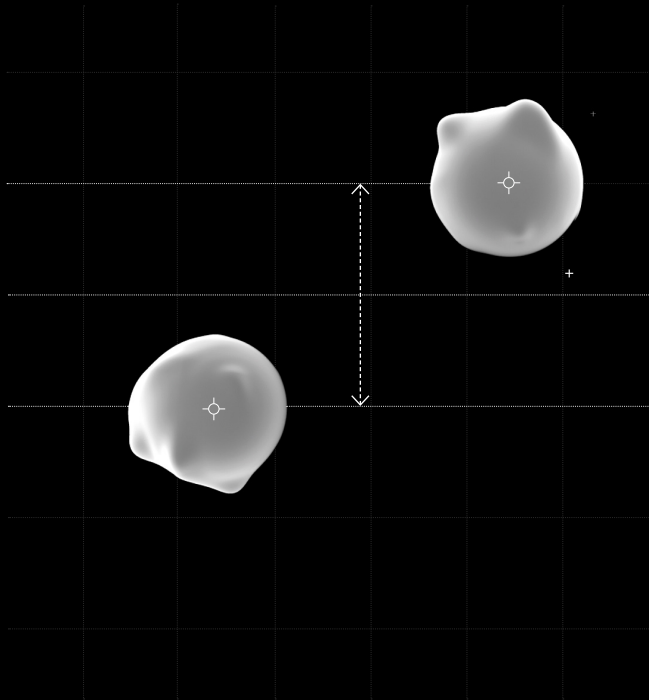
II. PARALANGUAGE

The simple, initially speech-generated blobs are further modified by a set of paralanguage parameters that encode two extra layers of detail and complexity. To achieve this, each shape’s texture is subjected to a Perlin noise algorithm that alters the object’s surface based on average pitch and average voice breaks. However, additional details can only be rendered retroactively, as both pitch and voice-break analyses imply averaging each speaker’s speech style in order to identify deviation from their ‘normal’. In this sense, pitch, measured in hertz,

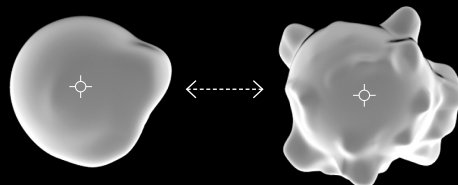
01 /
DURATION
(seconds)



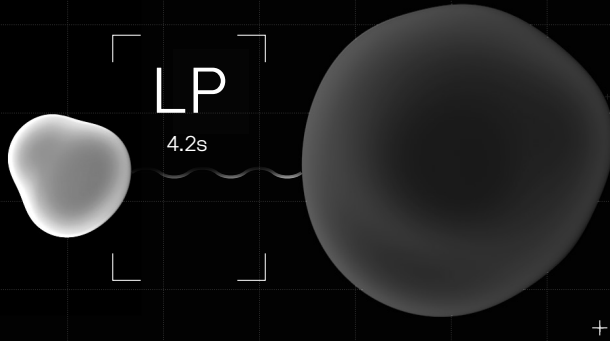
02 /
LOUDNESS
(decibels)



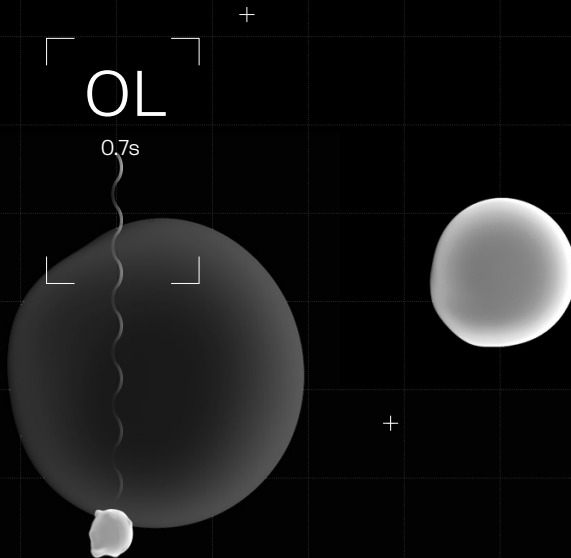
03 /
PITCH
(hertz)



04 /
LONG PAUSE
(> 3.0
seconds) +



05 /
OVERLAPS
(seconds)



is encoded in the density of the procedural texture (Figure 26, 3), with averagely uttered turns being rendered as soft-surfaced blobs, whereas marked deviations from average (seen as acts of exaggeration) are rendered as distressed, reflecting the overall deviation from the speaker's average.

Consequently, shape deformations are directly correlated to the supposed amount of speech exaggeration. Similarly, voice breaks, measured in percentage, are a marker of the overall mood and attitude of the speaker. They are directly relevant to the speed of an utterance and thus, they are naturally encoded in the texture's animation speed. In this sense, each bubble's speed of movement is inversely proportional to the overall amount of voice breaks recorded for that turn – with softly delivered turns (high percentage) being rendered with gently morphing texture, while quickly uttered turns (low percentage) rendered with rapidly evolving textures.

Further, another element of paralanguage (loudness as measured in decibels) is logically encoded in the bubble's position on the Y-axis, as an intuitive abstraction of the volume at which a sequence was uttered (Figure 26, 2). Similarly, loudness is only rendered retroactively as it is represented as a deviation from the overall average of the conversation.⁸

III. STRUCTURE

Lastly, at a 'zoomed out' level, patterns of an emerging social landscape may become discernible. By navigating the topography of the generated visualisations, viewers can classify relevant patterns not only in the form and distribution of shapes but also based on the empty spaces in between. These, taken together, encompass the impression of sequence design and interactional activity or lack thereof. The different kinds of spaces in between shapes, denoting the silences occurring throughout, build up to a different impression of what they might mean. To whom do these silences belong? Are they creating a jarring lapse in conversation? Are they ending the conversation altogether? Similarly, blobs occupying the same space denote overlaps that, too, can be interpreted in various ways: are they transitional overlaps (when speakers enter the conversation right

⁸ As such, there is first a need for analysing the mean loudness of an interaction, followed by an adjustment to the bubble's position in space to reflect their deviation from average, or zero on the Y-axis.

at the point of completion) or are they progressive overlaps (resulting from speech disfluency) (Figure 27)? Thus, by recognising the empty spaces in-between and around the shapes as a link between content and structure, viewers are provided with a richer topographic representation that blends into a single concept space able to convey the subtler aspects of the interaction.

Finally, these visualisations offer an element of interactivity, allowing the viewer to interact with each blob individually in order to reveal the extra connections and insights into the roles of particular utterances in the overall discourse. By interpreting the spatial relationships between blobs, their formation and detail, as well as signs or map annotations, markers of both harmony and disfluency become more apparent, pointing the viewer in the direction of tensions or incongruities. This is briefly demonstrated in the following three case study topographies, each involving a different AI agent performing mundane tasks.

D. FINDINGS

So, what does a Human-AI conversation look like? As it turns out in the following case studies, although the interactive tasks were similar, the final topology for each AI assistant, namely Siri, Alexa and Google Home, look quite distinct. There are, however, some key similarities. Most notably, there is a substantial imbalance in the overall contribution to the conversation, with the AI assistants visibly occupying more space than the users. This disproportion is far more evident in the interaction with Alexa and Siri than with Google Home, but is nonetheless present throughout. It should also be noted that very different interaction styles were observed between users, contributing to an overall impression of their individual approach and to the impression of what might constitute a sign of deviance or dissonance in each case:

SIRI

What becomes immediately visible from this topographical representation of a user asking Siri to recommend a movie is the noticeable disparity in the contribution to the interaction (Figure 28 & Figure 29). Looking at the first five turns [01], there is already a discernible pattern of disfluency with the user initiating an interaction that is not (visually) matched in style and content by the AI.

OL (0.7)

[03]

20.0

-10.0

OL (0.3)

[02]

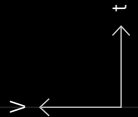
0.0

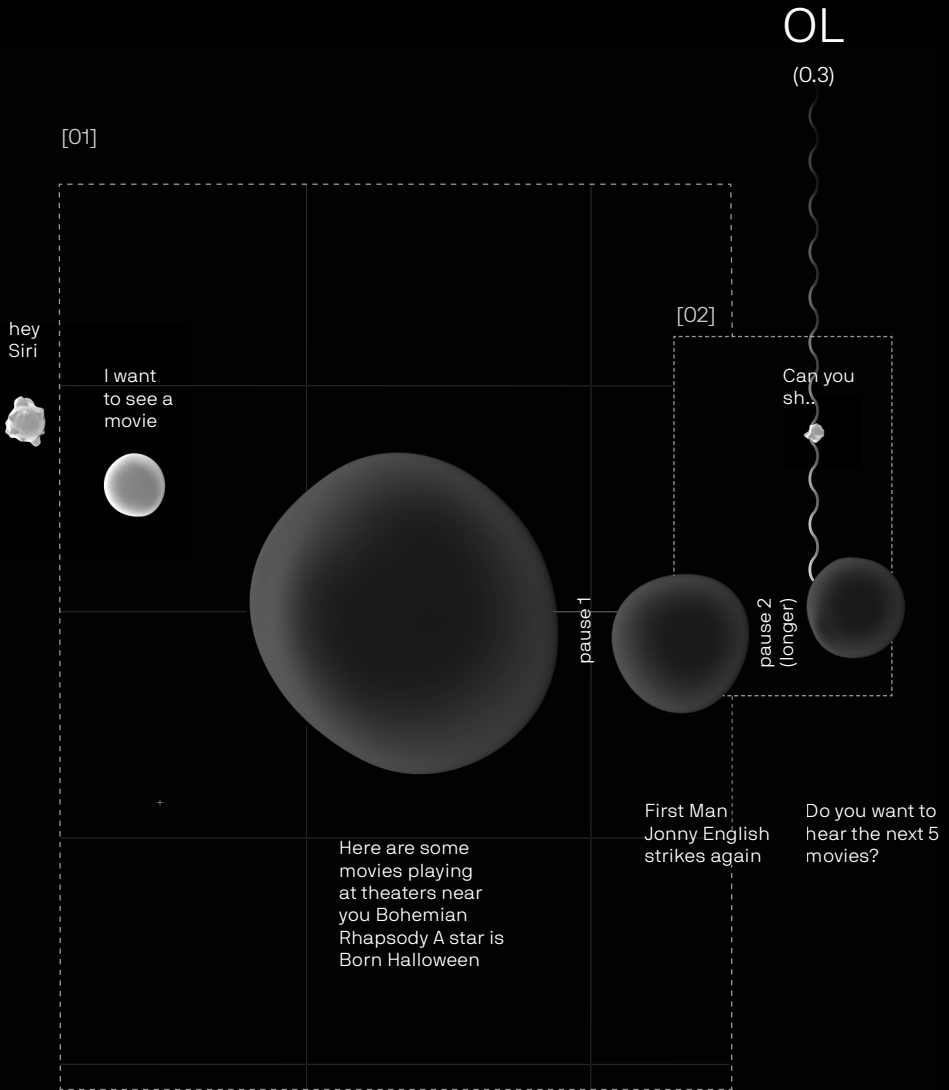
[01]

70

65

60





This disfluency, in turn, appears to lead to efforts towards correcting it, which can be seen by looking at the structural layer of the interaction, starting at the first overlap of blobs (Figure 29)[02]. It is noticeable that the user intervened at what they believed to be a point of completion, which, however, turned out to be a momentary false affordance. This overlap thus occurred due to the speech disfluency of the previous speaker, the AI, in failing to signify the type of silence intended or in failing to maintain an established interactional rhythm. Then, the resulting silence, perceived by the user as being longer-than-the-previous, was naturally interpreted as a cue for self-selection, leading to what could have been a preventable dissonance. Furthermore, the user's corrective attempt proved unsuccessful, with the agent continuing to deliver inconsistently sized responses with varying pauses in between, visually occupying almost the entire space of the conversation [03]. This inconsistency makes it very difficult, if not impossible, for the user to know when to intervene, as their mental model of what constitutes a pause or a silence keeps shifting in response to the AI's unpredictable speech style.

Therefore, this seemingly simple arrangement of shapes provides enough information to invite practical interpretations of both the interaction design flaws as well as users' erroneous mental models of it. A designer reviewing this visualisation might choose to resolve this dissonance by proposing affordances that more clearly communicate the difference between a pause in a series of turns and a silence planned to allow for the user to intervene – for instance, a visual representation of whether the silence 'belongs' to the AI agent or the user.

GOOGLE HOME

Here we notice a markedly different user interaction style, with quick turns scattered across both timeline and volume (Figure 30 & Figure 31). The user appears to fill in the silence with rapid utterances designed to occupy the pauses that 'belong' to them until they are able to select the next speaker – the AI agent. This could very well be a mark of a personal style, frequently occurring when speakers participate in the interaction enthusiastically, and exchange turns with continuity. However, the different speech styles appear to be problematic. What immediately attracts attention in this visual topography is the relatively small sizes of the user's blobs, scattered at very different points on the Y-axis, which stand in stark contrast to the AI's monotonous delivery [01][02]. Here, it is noticeable how the AI consistently delivers roughly same-length responses, with very little to

no paralanguage variety. This lack of variety could become frustrating for those who may wish to engage in a more humanlike interaction with their AI assistants, further conditioning the users to resort to interactions devoid of uniquely human cues. Thus, a designer studying this visualisation might choose to resolve this dissonance by proposing affordances that more clearly match the user's paralanguage style – for instance, instead of delivering one consistent paragraph of similar word-count, delivery could be split into multiple shorter ones that match the user's speech style. Next, by looking at the first distressed shape [03], delivered well below the average volume of the conversation (appearing to be a shape of a nervous laughter), it indicates that the user did not intend to address this utterance to the agent but rather to themselves. This could be seen as a sign of frustration or perhaps even tension in the conversation, denoting a conscious effort to attenuate or pass the situation along more swiftly. By digging deeper into the content of the shape that produced the tension (Figure 31) – “I searched the web and found this song oh yeah” – we notice how the apparent lack of context and unfortunate wording took the user by surprise, which indeed, led to a conscious effort to diminish the perceived clumsiness of the interaction.

ALEXA

This interaction topography of a user interacting with an Amazon Echo is riddled with issues comparable to those raised in both previous studies, yet noticeably further complicated by the several gaps and lapses present throughout (Figure 32 & Figure 33). The recurrent lapses – a type of silence defined as the failure of the current speaker to select the next speaker, causing the conversation to end, even if for just a moment – must be interpreted as an obvious sign of disfluency.

These long pauses [01][02] appear to suggest that the AI agent does not expect a response after their turn, which noticeably frustrates the user (Figure 33). As a result, we notice how visually, the user's blobs consistently alternate in size and detail, which, paired with the lack of variation on the Y-axis, suggest that the user is calmly attempting to synchronise the interaction model with their mental model. This attempt, while tense, nevertheless provides further insight into what the user judged as deficient in the previous interaction model. These repair strategies paint a stronger picture of the user's changing assumptions of the device's knowledge, while simultaneously highlighting the user's behavioural efforts to achieve coherence between their expectations and observed reality in resolving these series of momentary false affordances.

LP (3.8)

MUSIC

[02]

[01]

P (3.0)

[03]

30 Google Home Case Study Topography

20.0

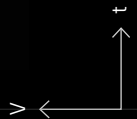
-10.0

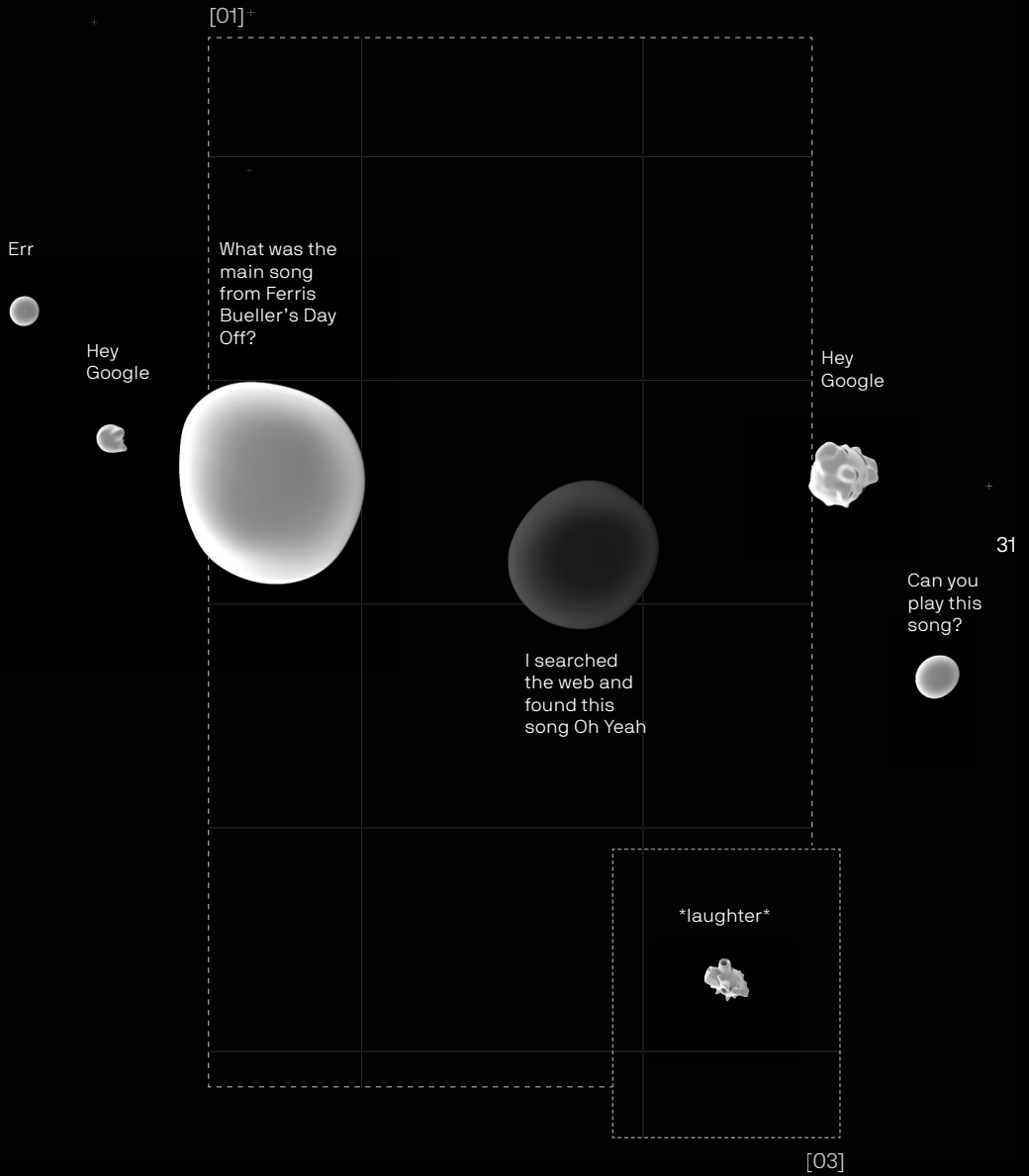
0.0

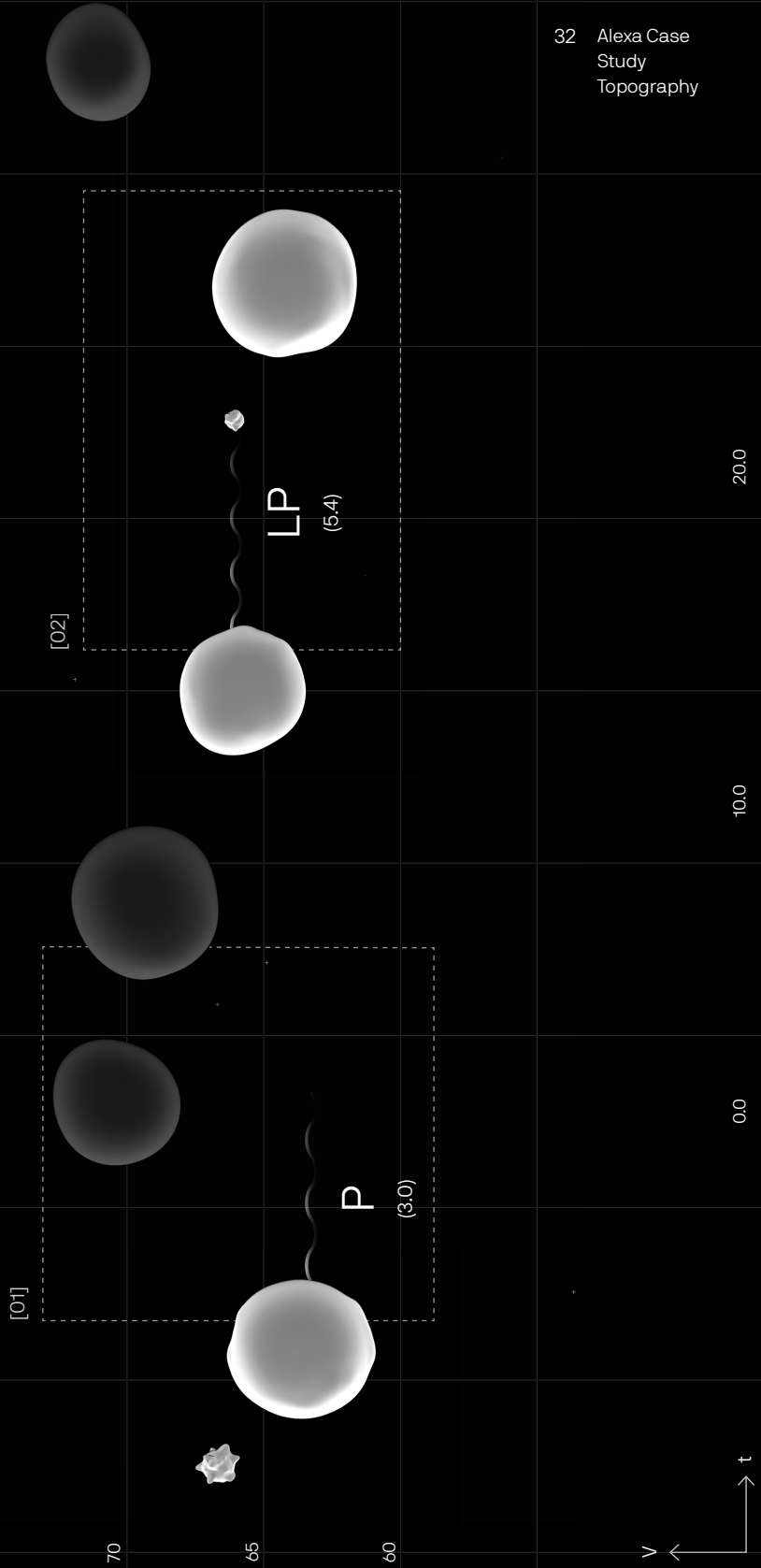
70

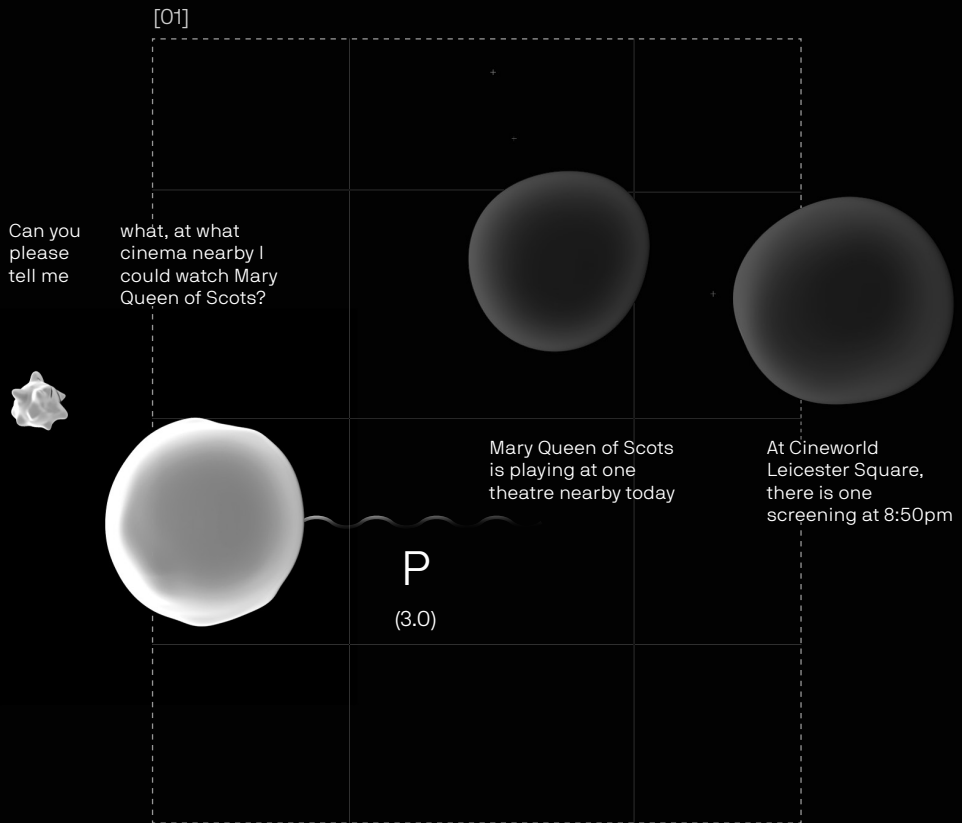
65

60









However, just as in the previous case study, these corrective strategies encompass valuable information about approaching a more fluent interaction design, one that is nevertheless in line with natural behavioural observations. In this instance, the user kept track of the task and continued narrowing down and further defining the options, unable to estimate the agent's knowledge of the topic at any given time. This could have been avoided by providing the means to successfully anticipate when the AI agent is proposing a gap rather than a lapse – a time in which the participant can self-select to advance the interaction rather than start a new one, with different assumptions of knowledge. This lack of short-term 'communication memory' is by no means a new problem. However, it now has a visual representation supporting the evidence of how it gives birth to dissonances in the Human-AI interaction, with further insights as to how might design practices tackle it more effectively.

This method unmistakably requires a degree of spontaneous interpretation, which aligns with the overarching philosophical perspective employed throughout this research. In identifying features of society or human affairs, designers ought to look for culturally-derived and contextually-situated interpretations of social life, a pursuit very much in line with a symbolic interactionist approach. As such, in the quest for discovering what regularly occurs, unnecessary definitions have been purposefully left out of the visualisation technique. The decision to avoid clear-cut interpretations was made as a means of avoiding the error of interpreting human behaviour as 'universal law'. In this vein, annotations are not interpretations but rather measurable elements of an interaction that need further investigation and personal analysis to discover what really occurs when studying unique aspects of a particular AI phenomenon.

E. DESIGNING ONWARDS

How has this investigation/methodology contributed, then, to the practice of designing technological interventions for social settings?

The *Just Like Me, But Not Exactly* set of experiments followed the prevailing approach to designing Human-AI interactions, one that does not require the researcher or the designer to locate themselves in the social world and to reflect

on how the simplest design choices contribute to the emergent Human-AI dynamic. It also does not call for including actual humans (as opposed to imagined humans represented by datasets) into the design process. Unsurprisingly then, design choices often have real social repercussions that go beyond what is anticipated (see privacy invasion, stereotype reinforcements etc.).

The methodological framework proposed here instead encourages the researcher or the AI designer to analyse the Human-AI dynamic as an emergent phenomenon riddled with personal inconsistencies and fluctuations that often go unnoticed. While not claiming to solve the issue of unanticipated social repercussions that may result from their design choices, it does give them a tool to look more closely at the interaction. The visual analysis developed in this chapter, though, is just an example of a potential method of studying what is locked away in the more subtle behavioural cues, arising moment by moment in the interaction – a method that facilitated my own reflection on the more transient and often chaotic dynamic between the Human and the AI.

04

CONCLUSIONS

OVERVIEW

The dichotomy between AI's technological and social facets has always been problematic, but perhaps even more so in the case of systems that, despite being intended to emulate sociality, are largely designed with a technology-driven mindset. Despite the need for systems more aligned with human expectations and values, it remains extremely difficult to computationally embed concepts as fundamentally fluid and situational as value and meaning. This is partly because present-day research is trying to formalise the principles of human behaviour, in all their complexity, through predictive machine learning models built on data extracted from how people behave, not *in relationship to AI* but, *in the absence of it*. As a result, the idea that the Human and the AI form an inseparable pair often remains unexplored.

In response, this research exposes the limitation of proceeding under the assumption that effects can be extracted from how humans behave in contrived situations, externally to any lived experiences or social contexts: first by exploring how people think of humanlike systems not in relation to them, but in the absence of them and secondly, by divorcing affordances from how humans process the perceptual information conveyed by the AI design.

As noted, however, the Human-AI phenomenon was not fundamentally stable, and the *imagined* humans for which these AI agents were designed were very different to the real-life, meaning-constructing humans who adjusted their behaviour and reasoning in response to the complex context of an interaction. With combinations of effects born in the fleeting concept-perception interaction, this work arrives at the conclusion that very few behavioural patterns and mental models can truly be universalised and one context cannot be neatly applied to the next.

As a result, this research proposes a shift in what constitutes an object of study – moving away from the object-based outputs of design practice, by arguing that for *any* output of a practice governed by an anthropomorphic frame the resultant object / interaction inherits the limiting assumptions of that frame. In its place, I am arguing for a reconfiguration at the level of the practice which then informs formal design decisions about the object.

In drawing together ideas from sociopsychology, HCI, HRI and design research, this thesis speaks to the necessity of creating a more fitting field of study, which implements the Human-AI object of study as a ground upon which to create new research. Therefore, this work does not address existing groups/communities of practice as they currently lack the transdisciplinary approach articulated as a result of my research findings. Instead, it encourages designers, technologists, researchers to adopt methodologies that better reflect the complexity of this nascent socio-technological world and its amended object of study.

In conclusion, in response to this investigation's main research question, '*How can design research navigate the gap between AI's social conceptualisations and human expectations?*', this study provides answers (to various extents) on three separate levels – theory, methodology and practice – as follows.

CONTRIBUTION TO KNOWLEDGE

A. THEORY

1. ANTHROPOMORPHISM IS MORE THAN A 'MINDLESS' SOCIAL REACTION.

Despite the pressing need to better understand people's tendency to anthropomorphise, very few experimentations have rigorously examined the mechanism behind this perceptive phenomenon. As shown, very little research on anthropomorphism tries to provide a cognitive justification for when and why people are likely to anthropomorphise nonhuman agents, instead the focus remains on examining the precision and functionality of these anthropomorphic portrayals. This often leads to the general belief that our reaction to humanlike design is merely a 'perceptive flaw'.

Contrary to the idea that people's responses to media are fundamentally social – reacting to different media as they would to other people, in the absence of an anthropomorphic model – this investigation has demonstrated that anthropomorphism is not a "mindless reaction" (see Reeves & Nass, 1996).

In doing so, it is answering this investigation's first sub-question: *Is human behaviour driven by mental heuristics independent of a cognitive theory or is it directed by more precise social schemas?*

Far from being independent from a cognitive theory, there are indeed precise social schemas observable in such interactions, as well as existing socio-psychological theory that can help explain the nature, consequences, and correlations of aspects of such interactions. This deceptively negligible concern has, however, important implications for the practice of design: the major difference between mindlessly enacting social habits and having a model of why a system behaves the way it does is that the former proposes automatic behavioural responses, in the absence of specific mental models, while the latter posits that a theory is the main drive behind people's attitudes. This implies that if the first explanation (mindless response to social cues) is correct, then designers should mainly focus on the social cues a system must emit to elicit desired user behaviour, whereas if the second explanation is true, design ought to primarily attend to the hierarchy of humanlike affordances that prime users' expectations of an interaction. This study has indicated that, at least in the context of humanlike AI designs, the latter is the appropriate approach.

2. ANTHROPOMORPHISM IS NOT PREDOMINANTLY OBSERVER-DEPENDENT.

Secondly, by borrowing from theories of cognition, another prevalent belief is that anthropomorphism, much like any other multifaceted cognitive process, is believed to be predominantly observer-dependent. Specifically, the literature on anthropomorphism as a perceptive phenomenon broadly predicts the variability of the phenomenon to be dictated by three distinct determinants – elicit agent knowledge, effectance and 'sociality' – each element making unique predictions about the strength and manifestation of the anthropomorphic perception.

Localising this gradient not on the external humanlike traits that inform these predilections but on personal, psychological determinants, appears to have important implications for Human-AI practice. Such an approach might lead researchers and designers to regard anthropomorphism as an inherently stable cognitive effect, which, as we have seen, is, at best, ill-defined. This investigation revealed a clear limitation in trying to predict anthropomorphism based purely on psychological determinants, and the three-factor theory performed poorly

in predicting participants' anthropomorphic gradient. Remarkably, theories on cognitive consistency, with their focus on beliefs as the building blocks of attitude structures, have proven to be a more suitable framework for interpreting anthropomorphic variability. This suggests that the anthropomorphic gradient could not have been easily extracted from human cognitive predispositions external to the interaction.

This work, then, argues for a reconfiguration at the level of the practice – away from an approach that readily embraces the anthropomorphic metaphor to one that more rigorously interrogates its inevitability.

3. AFFORDANCES CAN BE FLEETING AND CONSTRUCTED IN THE INTERACTION

The concept of affordance, while popular within the design community, remains ambiguous and with widely varying uses and definitions. The ambiguity can be traced back to its two main definitions, which on one side, sees affordances as only belonging to the environment (Gibson), while on the other side, sees them as tightly coupled with actor's past knowledge and experience (Norman).

These approaches have been discussed to be somewhat incomplete, as they bestow upon the concept of affordance a degree of constancy that is not necessarily reflected in practice. As we have seen in the experimentations on the AI artefact, false affordances, i.e., affordances that point towards a non-existent utility, may be the result of the unintended design of perceptual information (such as gender or accent), or may be the result of sequential inferences (such as inferring specialised knowledge from the interaction). As a result, what ends up becoming a false affordance is far less foreseeable as one signifier can lead to a series of other, perhaps unexpected elements of cognition, which can hardly ever be predicted in a practice governed by an anthropomorphic frame. This finding gives further weight to the necessity of curbing the design tendency that accords the artefact an elevated status to an emphasis on the Human-AI interaction in its place.

B. METHODOLOGY + METHOD

4. DISSONANCE METHODOLOGY

A consequence of the above-mentioned refashioned theoretical model of anthropomorphism was a shift in research strategy. By no longer seeing anthropomorphism as observer-dependent and instead subjected to fluctuations in relation to the interpretative processes of social interaction, the methodology employed in the first three experiments was abandoned in favour of one that required the AI designer to locate themselves in the emerging socio-technical world and to reflect on how their interventions impact the area for which they are making the observations.

In the first chapters, this work has also embraced the behaviouristic philosophy of pragmatism by adhering to the assertion that “the test of ideas, of thinking generally, is found in the consequences of the acts to which the ideas lead” (Dewey, 1929, p. 136). However, as we have seen in the first two perspectives on Human and AI, the resulting actions offered only a limited analysis of the dynamic between the conceptualisation and perception. In both cases, to account for the observed phenomena, I had to eventually resort to how participants interpreted, acted and interacted as a result of that understanding, in order to account for the resulting behaviour. As a result, this study ultimately embraced the symbolic interactionist perspective that “human beings respond to a particular situation through how they define that situation, rather than how the situation is objectively presented to them” (Aldiabat and Navenec, 2011, p. 1067), subsequently shifting the emphasis from the resulting action to the construction of meaning.

Consequently, in contributing to how we research and design AI and in answering this investigation’s third sub-question – *how are expectations continuously constituted through the interaction between the two, in practice?* – I proposed weaving symbolic interactionism, with its emphasis on meaning, interaction and interpretation into a new design research strategy. As such, the resultant Dissonance Methodology, refashioned by reclaiming its potent tools of analysis from their time-consuming and dated origins (Dissonance Research) has been proposed as an open-ended methodology with an inherent emphasis on the interpretative process and the resulting action.

Furthermore, this framework which sees *inconsistency* as a motivating factor – exercising a psychologically uncomfortable *tension* that motivates individuals to achieve consonance through various strategies (*elimination, reduction or avoidance*) – further unveils the histories and dynamics of the co-construction of meaning. This middle ground between the two strands (resulting action and interpreted meaning) proved to be a favourable way of investigating naturally occurring interactions between the Human and AI in both research and practice.

However, this shifting perspective posed its own challenges, one being the lack of available methods suited to grapple with the complexity of this emergent social world between the Human and the AI. In addressing this, I have also proposed a method for visualising tension as locked away in the more subtle behavioural cues that arise moment by moment in the interaction.

5. VISUAL ANALYSIS METHOD

In line with the above-mentioned methodology, I have further explored a digital method of encoding inconsistencies visually, by mapping linguistic productions onto a generative, topographical landscape that could aid in the interpretative process. As a result, the planned method considered the entire conversation as a unit of analysis, looking not only at what is said, by whom it is said and for what purpose, but also in what way the meaning of the text is influenced by speech styles and context.

The visual method breaks down Human-AI interactions into components, starting from one turn as a unit of analysis and then gradually constructing an artificial conversational landscape that encrypts further layers of social complexity. However, in staying true to the very specific flavour of interpretative meaning, the analysis still requires an unmistakable degree of spontaneous analysis, which is in line with the overarching philosophical perspective employed throughout this research. It, therefore, does not seek to confer clear evidence of inconsistencies or strategies, but rather only to facilitate the process of identifying outliers and looking for culturally derived and contextually situated interpretations of social life.

Therefore, the repurposed methodology and method contribute to a development in how design and research may explore the layers of social complexity emerging from our interaction with AI systems and how to address, in practice, some of the more fleeting aspects of this interaction.

C. DESIGN PRACTICE

6. DESIGNERS' ROLE

Finally, this research contributed towards the practice of designing humanlike AI agents first by reviewing designers' roles in relation to eliciting anthropomorphic mental models. Namely, the practice's fondness of viewing anthropomorphism as dictated by personal, psychological determinants renders anthropomorphism as an inherently stable cognitive effect, which is, at best, an incomplete model. By failing to see it as fluctuating in response to minimal perceptual cues, designers may rely on people's intrinsic predilections for humanlike inductions and a presumed natural process of cognitive correction, which, as seen in the first experiment, is flawed.

It was also noted that anthropomorphic mental models are incredibly difficult to correct and are further sustained by highly humanlike interactions paired with a deep-seated technological obscurity. This argument was substantiated in the first experiment through the discovery that, even with ample space for correction, users' unreliable humanlike mental models are carried over time with little to no correction. The evidence suggests that any tensions between users' schemas and the nature of the interaction would not be alleviated on their own, and designers of personified forms must account for these dissonances 'upstream' in the design process, instead of so easily resorting to 'indistinguishability' as an incentive. Anthropomorphism may be inevitable, but it is not unproblematic. Failing to account for this complexity in practice is a delegation of responsibility.

7. FALSE AFFORDANCES ARE PROBLEMATIC

In the *Just Like Me, But Not Exactly* series of experiments, while answering this investigation's second sub-question – *are anthropomorphic expectations considerably different from their human-reality counterparts?* – it was noted that a very similar, if not identical mechanism through which we make inferences of personified AI agents existed. Because of this, the presence of controlled perceptual information leads not only to correct perceptible affordances (for example, the ability to 'understand' English phrases) but also to other apparent affordances that might not map to an actual utility (knowledge of landmarks). This grouping of information most notably results in an inability to correctly reject false

affordances, contributing to the argument against personification in design (as inescapably leading to the false expectations of erroneous mental models).

8. HIERARCHY OF AFFORDANCES

This lumping of affordances appears to be the result of a well-known mental heuristic – the fundamental attribution error – implying that signifiers pertaining to the internal properties of an agent (such as personality) will be preferred over those pertaining to external (technical or form-based) properties, despite the obvious absence of any ‘internal’ traits. The indication that much of users’ interaction results in a lumping of affordances because of a mental heuristic most notably denotes a necessity to first attend to the more salient ‘internal’ signifiers, as users’ expectations of a system are heavily influenced by their opinions of a particular persona (stereotypes). However, this goes both ways.

9. PERSONIFYING AI AGENTS MAY HAVE REAL-WORLD IMPLICATIONS

The fact that participants’ attitudes towards the device were principally influenced by their expectations of a distinct persona represents a powerful illustration of how projecting digital assistants as belonging to distinct social groups could inevitably reinforce potentially harmful stereotypes. The mere association between the device’s function and its persona unavoidably invites the plethora of expectations associated with both the function and the ‘type of person’ it portrays, and thus, regardless of whether it meets users’ expectations or not, it still conditions people to fall back upon pre-existing schemas.

10. MORE HUMANLIKE IS NOT BETTER

While AI design seems to be profiting from eliciting anthropomorphic beliefs, there is quite clearly no sharp border imposed on the strength of these beliefs, leading to an unwritten conviction that the more humanlike an interaction appears to be, the higher the ease with which users interact with it. It was noted, however, that this correlation of events is not so straightforward and users’ attitudes are perhaps better justified by shifting perspective to account for how participants responded to the situation based on their interpretation of that situation.

11. LESS HUMANLIKE IS NOT BETTER

While interactions with AI assistants see people heavily rely on the same cognitive tools utilised in predicting other people's behaviours, the mechanism does not perfectly overlap with our expectations of humanlike AI agents. In certain situations, such as entertainment tasks, different schemas may be at play, and approaching AI humanlike design from a less literal emulation of human-to-human interaction does not necessarily address our biggest concerns associated with these interactions. In practice, it was noted how diluting humanlike design traits might alleviate some problematic aspects of these interactions, such as trust, but generate other false affordances in the process. These findings indicate that, in designing humanlike AI agents, there may be no straightforward recipe since what is expected and appropriate when designing humanlike agents persistently fluctuates in response to the overabundance of internal cues and the external signifiers of the interaction.

FUTURE RESEARCH

As this exploration ultimately implies, there are still many aspects of the Human-AI dynamic left to be uncovered. It is, therefore, important to consider how to address these newly created questions. First, to some extent, the introduction of dynamic processes or systems within the practice of designing AI appears to come into conflict with the industry's affinity with efficiency. By definition dynamic systems, such as the Human-AI model posited in this thesis, are models that: behave in a nonlinear manner, are extremely sensitive to minor changes and usually give rise to completely unpredictable behaviours. Acknowledging this in practice may ultimately seem daunting, and further research is needed to make it more accessible.

One potential direction moving forward would be to draw analogies with other dynamic systems in which, although outcomes are chaotic and unpredictable for any given moment in time, the accumulation of chaos and unpredictability ultimately creates coherent patterns. Some psychologists have already begun to apply the framework of chaos and complexity to human behaviour (Guastello, Koopmans and Pincus, 2009) and there is strong evidence to indicate that cognitions tend to self-organise in coherent patterns over time (Read and

Simon, 2012). Further work is needed to explore how patterns of behaviours, as they emerge over time in the interaction with AI (rather than in isolated events), may contribute to our understanding of this socio-technological landscape and towards what is expected and what is unique to particular situations.

Building on this idea, generating a catalogue of different ‘inconsistencies’ would advance not only the theoretical necessity of clarifying the term, but also our understanding of how dissonance may vary from one context to the next. As a result, a dictionary of ‘patterns’ might facilitate the design process in its ability to predict how particular dynamics are more likely than others to occur, depending on the context of the Human-AI interaction.

Developing a theory of such a nonlinear phenomenon, accounting for the chaos and complexity of the Human-AI dynamic, certainly requires further research.

I began this research with a clear identification of a knowledge gap between the social and technological facets of AI but with no blueprint for how to tackle what turned out to be a vastly more complex Human-AI mutuality than initially anticipated. In navigating the concept-percept phenomenon, I gradually arrived at the conclusion that predominant AI research and design practices were falling short on addressing some of the more fleeting aspects of the interaction. In my view, the predominant approach was not naturally fitting within this subject of study. While retaining the perspectives of AI as a cultural artefact and anthropomorphism as an emerging phenomenon, the central research question guiding this study has led to the proposal and initial development of a new model better suited to navigating this dynamic, nonlinear, emergent field of study: the Human-AI.

()
**LIST OF
REFERENCES**

- Aldiabat, K. and Navenec, C.-L.L. (2011) 'Philosophical Roots of Classical Grounded Theory: Its Foundations in Symbolic Interactionism', *The Qualitative Report*, 16(4), pp. 1063–1080. doi:10.46743/2160-3715/2011.1121.
- Amazon.com, I. (2020) *Introducing the All-New Echo Family—Reimagined, Inside and Out* | Amazon.com, Inc. - Press Room, Amazon Press release. Available at: <https://press.aboutamazon.com/news-releases/news-release-details/introducing-all-new-echo-family-reimagined-inside-and-out/> (Accessed: 31 December 2021).
- Amershi, S. et al. (2019) 'Guidelines for human-AI interaction', in *Proceedings of the 2019 chi conference on human factors in computing systems*, pp. 1–13.
- 'anthropomorphism, n.' (1885) *Oxford English Dictionary*. 1st edn. Oxford: Oxford University Press.
- Aronson, E. (1969) 'The Theory of Cognitive Dissonance: A Current Perspective' This article and all the original work reported in it were supported by a grant from the National Science Foundation.', in Berkowitz, L. (ed.) *Advances in Experimental Social Psychology*. Academic Press, pp. 1–34. doi:10.1016/S0065-2601(08)60075-1.
- Aronson, E. (1992) 'The Return Of The Repressed: Dissonance Theory Makes A Comeback', *Psychological Inquiry*, 3(4), pp. 303–311. doi:10.1207/s15327965pli0304_1.
- Austermann, A. et al. (2010) 'How do users interact with a pet-robot and a humanoid', in *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems - CHI EA '10. the 28th of the international conference extended abstracts*, Atlanta, Georgia, USA: ACM Press, p. 3727. doi:10.1145/1753846.1754046.
- Awad, E. et al. (2018) 'The Moral Machine experiment', *Nature*, 563(7729), pp. 59–64. doi:10.1038/s41586-018-0637-6.
- Baecker, R.M. and Buxton, W.A.S. (eds) (1987) *Readings in human-computer interaction: a multidisciplinary approach*. Los Altos, Calif: M. Kaufmann.

()

Bahník, Š., English, B. and Strack, F. (2017) 'Anchoring effect', in *Cognitive illusions: Intriguing phenomena in thinking, judgment and memory*, 2nd ed. New York, NY, US: Routledge/Taylor & Francis Group, pp. 223–241.

Bartneck, C. *et al.* (2005) 'Robot Abuse – A Limitation of the Media Equation', p. 4.

Bartneck, C. *et al.* (2007) 'Is The Uncanny Valley An Uncanny Cliff?', in *RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication. RO-MAN 2007 - The 16th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 368–373. doi:10.1109/ROMAN.2007.4415111.

Bartneck, C. and Okada, M. (2001) 'Robotic user interfaces'.

Bethel, C.L., Salomon, K. and Murphy, R.R. (2009) 'Preliminary results: humans find emotive non-anthropomorphic robots more calming', in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction - HRI '09. the 4th ACM/IEEE international conference*, La Jolla, California, USA: ACM Press, p. 291. doi:10.1145/1514095.1514181.

Bloomfield, B.P. (ed.) (2018) *The Question of Artificial Intelligence: Philosophical and Sociological Perspectives*. 1st edition. Routledge.

Blumer, H. (1986) *Symbolic Interactionism: Perspective and Method*.

Boersma, P. and van Heuven, V. (2001) 'Speak and unSpeak with PRAAT', 5(9), p. 7.

Bolukbasi, T. *et al.* (2016) 'Quantifying and Reducing Stereotypes in Word Embeddings', *arXiv:1606.06121 [cs, stat]* [Preprint]. Available at: <http://arxiv.org/abs/1606.06121> (Accessed: 23 December 2021).

Boyd, R. and Holton, R.J. (2017) 'Technology, innovation, employment and power: Does robotics and artificial intelligence really mean social transformation?', *Journal of Sociology*, 54(3), pp. 331–345. doi:10.1177/1440783317726591.

Boyer, P. (1996) 'What makes anthropomorphism natural: Intuitive ontology and cultural representations', *Journal of the Royal Anthropological Institute*, pp. 83–97.

Brehm, J.W. and Cohen, A.R. (1962) *Explorations in cognitive dissonance*. - *PsycNET*. Available at: <https://psycnet.apa.org/doiLanding?doi=10.1037%2F11622-000> (Accessed: 20 February 2022).

Brunswik, E. (1956) *Perception and the representative design of psychological experiments, 2nd ed.* Berkeley, CA, US: University of California Press (Perception and the representative design of psychological experiments, 2nd ed).

Bryant, A. (2002) 'Re-Grounding Grounded Theory', in.

Bunge, M. (1963) 'A General Black Box Theory', *Philosophy of Science*, 30(4), pp. 346–358. doi:10.1086/287954.

Callon, M. (2005) 'Peripheral Vision', *Organization Studies - ORGAN STUD*, 26, pp. 1229–1250. doi:10.1177/0170840605056393.

Caporael, L.R. (1986) 'Anthropomorphism and mechanomorphism: Two faces of the human machine', *Computers in human behavior*, 2(3), pp. 215–234.

Carbon, C.-C. (2014) 'Understanding human perception by human-made illusions', *Frontiers in Human Neuroscience*, 8. doi:10.3389/fnhum.2014.00566.

Carley, K. and Palmquist, M. (1992) 'Extracting, representing, and analyzing mental models', *Social forces*, 70(3), pp. 601–636.

Carmon, Z. et al. (2019) 'Designing AI systems that customers won't hate', *MIT Sloan Management Review* [Preprint].

Castelli, F. et al. (2000) 'Movement and mind: a functional imaging study of perception and interpretation of complex intentional movement patterns', *NeuroImage*, 12(3), pp. 314–325. doi:10.1006/nimg.2000.0612.

Cave, S., Dihal, K. and Dillon, S. (eds) (2020) *AI Narratives: A History of Imaginative Thinking about Intelligent Machines*. Oxford: Oxford University Press. doi:10.1093/oso/9780198846666.001.0001.

Chalmers, M. and Galani, A. (2004) 'Seamful interweaving: heterogeneity

()

in the theory and design of interactive systems', in *Proceedings of the 5th conference on Designing interactive systems: processes, practices, methods, and techniques*. New York, NY, USA: Association for Computing Machinery (DIS '04), pp. 243–252. Available at: <https://doi.org/10.1145/1013115.1013149>.

Charles M. A., C. and Aleksandr V., G. (2020) *Artificial Intelligence and Human Flourishing - Clark - 2020 - American Journal of Economics and Sociology - Wiley Online Library*. Available at: <https://onlinelibrary.wiley.com/doi/10.1111/ajes.12356> (Accessed: 23 December 2021).

Charmaz, K. (2000) *Grounded theory: objectivist and constructivist methods*. Thousand Oaks, Calif.: Sage.

Charmaz, K. (2013) *Constructing Grounded Theory*. Second edition. London ; Thousand Oaks, Calif: SAGE Publications Ltd.

Chomsky, N. (2002) *Syntactic Structures*. Walter de Gruyter.

Chun Tie, Y., Birks, M. and Francis, K. (2019) 'Grounded theory research: A design framework for novice researchers', *SAGE Open Medicine*, 7, p. 205031211882292. doi:10.1177/2050312118822927.

Collins, H. (1992) *Artificial Experts: Social Knowledge and Intelligent Machines*. Edited by W.E. Bijker, W.B. Carlson, and T. Pinch. Cambridge, Mass: The MIT Press.

Collins, H. (2018) *Artificial Intelligence: Against Humanity's Surrender to Computers*. 1st edition. Polity.

Consumer Intelligence Research Partners, LLC. (2017) 'Home Automation Device Market Grows Briskly to 27 Million'. Available at: <https://voicebot.ai/wp-content/uploads/2017/11/cirp-news-release-2017-11-06-echo-home.pdf>.

Cooper, J. (2007) *Cognitive Dissonance: 50 Years of a Classic Theory*. SAGE.

Corbin, J.M. and Strauss, A. (1990) 'Grounded theory research: Procedures,

- canons, and evaluative criteria', *Qualitative Sociology*, 13(1), pp. 3–21. doi:10.1007/BF00988593.
- Cowls, J. *et al.* (2019) 'Designing AI for social good: Seven essential factors', *Available at SSRN 3388669* [Preprint].
- Cross, E.S., Hortensius, R. and Wykowska, A. (2019) 'From social brains to social robots: applying neurocognitive insights to human–robot interaction', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771), p. 20180024. doi:10.1098/rstb.2018.0024.
- Crowelly, C.R. *et al.* (2009) 'Gendered voice and robot entities: Perceptions and reactions of male and female subjects', in *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 3735–3741. doi:10.1109/IROS.2009.5354204.
- Dahlin, E. (2021) 'Mind the gap! On the future of AI research', *Humanities and Social Sciences Communications*, 8(1), pp. 1–4. doi:10.1057/s41599-021-00750-9.
- Dastin, J. (2018) 'Amazon scraps secret AI recruiting tool that showed bias against women', *Reuters*, 10 October. Available at: <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G> (Accessed: 23 December 2021).
- Dennett, D.C. (1998) *The intentional stance*. 7. printing. Cambridge, Mass.: MIT Press (A Bradford book).
- Dewey, J. (1929) *The Quest for Certainty: A Study of the Relation of Knowledge and Action*. New York, NY: Putnam.
- Dignum, V. (2017) 'Responsible artificial intelligence: designing AI for human values'.
- DiSalvo, C. and Gemperle, F. (2003) 'From Seduction to Fulfillment: The Use of Anthropomorphic Form in Design', *Proceeding DPPI '03 Proceedings of the 2003 international conference on Designing pleasurable products and interfaces*, pp. 67–72. doi:10.1145/782896.782913.

()

- Don, A. et al. (1992) 'Anthropomorphism: from Eliza to Terminator 2', in *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '92. the SIGCHI conference*, Monterey, California, United States: ACM Press, pp. 67–70. doi:10.1145/142750.142760.
- Donath, J., Karahalios, K. and Viégas, F. (1999) 'Visualizing Conversation', *Journal of Computer-Mediated Communication*, 4(4), pp. 0–0. doi:https://doi.org/10.1111/j.1083-6101.1999.tb00107.x.
- Duffy, B. (2003) 'Anthropomorphism and the social robot', *Robotics and Autonomous Systems*, 42, pp. 177–190. doi:10.1016/S0921-8890(02)00374-3.
- Duffy, B.R. (2002) 'Anthropomorphism and robotics', *The society for the study of artificial intelligence and the simulation of behaviour*, 20.
- Duffy, B.R. and Zawieska, K. (2012) 'Suspension of disbelief in social robotics', in *2012 IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication. 2012 RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, Paris, France: IEEE, pp. 484–489. doi:10.1109/ROMAN.2012.6343798.
- Epley, N. et al. (2004) 'Perspective taking as egocentric anchoring and adjustment', *Journal of Personality and Social Psychology*, 87, pp. 327–339.
- Epley, N., Waytz, A. and Cacioppo, J.T. (2007) 'On seeing human: A three-factor theory of anthropomorphism.', *Psychological Review*, 114(4), pp. 864–886. doi:10.1037/0033-295X.114.4.864.
- Etzioni, A. and Etzioni, O. (2016) 'Designing AI systems that obey our laws and values', *Communications of the ACM*, 59(9), pp. 29–31.
- Eyssel, F. et al. (2012) "'If you sound like me, you must be more human": on the interplay of robot and user features on human-robot acceptance and anthropomorphism', in *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction - HRI '12. the seventh annual ACM/IEEE international conference*, Boston, Massachusetts, USA: ACM Press, p. 125. doi:10.1145/2157689.2157717.

Fantini van Ditmar, D. (2016) *IdIoT: second-order cybernetics in the 'smart' home*. Ph.D. Royal College of Art. Available at: <http://researchonline.rca.ac.uk/2697/> (Accessed: 2 September 2019).

Festinger, L. (1957) *A Theory of Cognitive Dissonance*. Stanford University Press.

Fineman, B. (2004) 'Computers as people: human interaction metaphors in human-computer interaction', p. 52.

Fink, J. (2012) 'Anthropomorphism and human likeness in the design of robots and human-robot interaction', in *International Conference on Social Robotics*. Springer, pp. 199–208.

Floridi, L. *et al.* (2021) 'How to design AI for social good: Seven essential factors', in *Ethics, Governance, and Policies in Artificial Intelligence*. Springer, pp. 125–151.

Forsythe, D.E. (1993a) 'Engineering Knowledge: The Construction of Knowledge in Artificial Intelligence', *Social Studies of Science*, 23(3), pp. 445–477.

Forsythe, D.E. (1993b) 'The Construction of Work in Artificial Intelligence', *Science, Technology, & Human Values*, 18(4), pp. 460–479.

Freud, S. (1976) *The uncanny*. Penguin London.

Frey, C.B. and Osborne, M.A. (2013) 'The future of employment: How susceptible are jobs to computerisation?', *Technological Forecasting and Social Change*, 114, pp. 254–280. doi:10.1016/j.techfore.2016.08.019.

Gaver, W.W. (1991) 'Technology affordances', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '91), pp. 79–84. doi:10.1145/108844.108856.

Gaver, W.W. (1996) 'Situating action II: Affordances for interaction: The social is material for design', *Ecological psychology*, 8(2), pp. 111–129.

Gentner, D. and Stevens, A.L. (eds) (1983) *Mental models*. Hillsdale, N.J: Erlbaum

()

(Cognitive science).

- Ghezeljeh, T.N. and Emami, A. (2009) 'Grounded theory: methodology and philosophical perspective', *Nurse Researcher*, 17(1), pp. 15–23.
- Gibson, J.J. (2014) *The Ecological Approach to Visual Perception: Classic Edition*. 1st edition. New York London: Psychology Press.
- Glanville, R. (1982) 'Inside every white box there are two black boxes trying to get out', *Behavioral Science*, 27(1), pp. 1–11. doi:10.1002/bs.3830270102.
- Glaser, B.G. (2002) 'Constructivist Grounded Theory?', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 3(3). doi:10.17169/fqs-3.3.825.
- Glaser, B.G. and Holton, J. (2004) 'Remodeling Grounded Theory', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 5(2). doi:10.17169/fqs-5.2.607.
- Gong, L. (2008) 'How social is social responses to computers? The function of the degree of anthropomorphism in computer representations', *Computers in Human Behavior*, 24(4), pp. 1494–1509.
- Gray, H.M., Gray, K. and Wegner, D.M. (2007) 'Dimensions of Mind Perception', *Science*, 315(5812), pp. 619–619. doi:10.1126/science.1134475.
- Guastello, S.J., Koopmans, M. and Pincus, D. (eds) (2009) *Chaos and complexity in psychology: the theory of nonlinear dynamical systems*. Cambridge ; New York: Cambridge University Press.
- Guba, E. and Lincoln, Y. (1994) 'Competing paradigms in qualitative research.', *undefined* [Preprint]. Available at: <https://www.semanticscholar.org/paper/Competing-paradigms-in-qualitative-research.-Guba-Lincoln/f4ee6f7b09f4b1c9943cc36a8aa5a6391e1a92cf> (Accessed: 24 November 2021).
- Guthrie, S.E. (1995) *Faces in the Clouds: A New Theory of Religion*. Oxford University Press.

- Guthrie, S.E. (1997) 'Anthropomorphism: A definition and a theory.'
- Harper, R.H.R. (2012) *Texture: Human Expression in the Age of Communications Overload*. Reprint edition. Cambridge, Mass.; London: MIT Press.
- Harrison, C. and Romhild, C. (2008) *Visualizing The Bible, Visual Complexity*. Available at: <http://www.visualcomplexity.com/vc/project.cfm?id=593> (Accessed: 20 February 2022).
- Harter, T. et al. (2012) 'A file is not a file: Understanding the I/O behavior of Apple desktop applications', *ACM Transactions on Computer Systems (TOCS)*, 30(3), pp. 1–39.
- Haslam, N. et al. (2005) 'More human than you: attributing humanness to self and others', *Journal of Personality and Social Psychology*, 89(6), pp. 937–950. doi:10.1037/0022-3514.89.6.937.
- Haslam, N. (2006) 'Dehumanization: An Integrative Review', *Personality and Social Psychology Review*, 10(3), pp. 252–264. doi:10.1207/s15327957pspr1003_4.
- Haslam, N., Bastian, B. and Bissett, M. (2004) 'Essentialist beliefs about personality and their implications', *Personality & Social Psychology Bulletin*, 30(12), pp. 1661–1673. doi:10.1177/0146167204271182.
- Havre, S., Hetzler, B. and Nowell, L. (2000) 'ThemeRiver: visualizing theme changes over time', in *IEEE Symposium on Information Visualization 2000. INFOVIS 2000. Proceedings. IEEE Visualization 2000*, Salt Lake City, UT, USA: IEEE Comput. Soc, pp. 115–123. doi:10.1109/INFVIS.2000.885098.
- Hayes, P. and Ford, K. (1995) 'Turing Test considered harmful', in *In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, pp. 972–977.
- Hegel, F. et al. (2011) 'Towards a typology of meaningful signals and cues in social robotics', in *2011 RO-MAN. 2011 RO-MAN: The 20th IEEE International Symposium on Robot and Human Interactive Communication*, Atlanta, GA, USA: IEEE, pp. 72–78. doi:10.1109/ROMAN.2011.6005246.

()

- Heider, F. (1958) *The psychology of interpersonal relations*. Hoboken, NJ, US: John Wiley & Sons Inc (The psychology of interpersonal relations). doi:10.1037/10628-000.
- Heider, F. and Simmel, M. (1944) 'An experimental study of apparent behavior', *The American Journal of Psychology*, 57, pp. 243–259. doi:10.2307/1416950.
- Higgins, E.T. (1989) *Self-Discrepancy Theory: What Patterns of Self-Beliefs Cause People to Suffer?* - ScienceDirect. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0065260108603068> (Accessed: 20 February 2022).
- Higgins, E.T. (1996) 'Knowledge activation: Accessibility, applicability, and salience', in *Social psychology: Handbook of basic principles*. New York, NY, US: Guilford Press, pp. 133–168.
- Ho, C.-C. and MacDorman, K.F. (2010) 'Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices', *Computers in Human Behavior*, 26(6), pp. 1508–1518. doi:10.1016/j.chb.2010.05.015.
- Hogg, P.M. and Vaughan, P.G. (2017) *Social Psychology*. 8 edition. New York: Pearson.
- Iacoboni, M. et al. (2004) 'Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline', *NeuroImage*, 21(3), pp. 1167–1173. doi:10.1016/j.neuroimage.2003.11.013.
- Jenkins, H.S. (2008) 'Gibson's "Affordances": Evolution of a Pivotal Concept', p. 12.
- Johnson, R. and Walsh, I. (2019) 'Mixed grounded theory: Merging grounded theory with mixed methods and multimethod research.', in.
- Johnson, R.B., McGowan, M.W. and Turner, L.A. (2010) 'Grounded Theory in Practice: Is It Inherently a Mixed Method?', *Research in the Schools*, 17(2), pp. 65–78.

- Kaaronen, R.O. (2018) 'A Theory of Predictive Dissonance: Predictive Processing Presents a New Take on Cognitive Dissonance', *Frontiers in Psychology*, 9, p. 2218. doi:10.3389/fpsyg.2018.02218.
- Kanda, T. et al. (2005) 'Analysis of humanoid appearances in human-robot interaction', in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems. 2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Edmonton, Alta., Canada: IEEE, pp. 899–906. doi:10.1109/IROS.2005.1544960.
- Kaptelinin, V. (2014) *Affordances and design*. Interaction design foundation.
- Karpinska-Krakowiak, M. and Eisend, M. (2021) *The Effects of Animistic Thinking, Animistic Cues, and Superstitions on Brand Responses on Social Media - ScienceDirect*. Available at: <https://www.sciencedirect.com/science/article/pii/S1094996821000189?via%3Dihub> (Accessed: 13 February 2022).
- Kelle, U. (2005) "'Emergence" vs. "Forcing" of Empirical Data? A Crucial Problem of "Grounded Theory" Reconsidered', *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research*, 6(2). doi:10.17169/fqs-6.2.467.
- Kennedy, J.S. (1992) *The new anthropomorphism*. Cambridge University Press.
- Kenny, M. and Fourie, R. (2015) 'Contrasting Classic, Straussian, and Constructivist Grounded Theory: Methodological and Philosophical Conflicts', *The Qualitative Report* [Preprint]. doi:10.46743/2160-3715/2015.2251.
- Kiesler, S.B. and Goetz, J. (2001) 'Mental Models and Cooperation with Robotic Assistants', in *CHI 2001*.
- Kim, Y. and Sundar, S.S. (2012) 'Anthropomorphism of computers: Is it mindful or mindless?', *Computers in Human Behavior*, 28(1), pp. 241–250. doi:10.1016/j.chb.2011.09.006.
- King, W.J. and Ohya, J. (1996) 'The representation of agents: Anthropomorphism, agency, and intelligence', in *Conference Companion on Human Factors in Computing Systems*, pp. 289–290.

()

Krementsov, N. and Todes, D. (2010) 'On Metaphors, Animals, and Us', *Journal of Social Issues*, 47, pp. 67–81. doi:10.1111/j.1540-4560.1991.tb01823.x.

Kunda, Z. (1990) *The case for motivated reasoning* - PubMed. Available at: <https://pubmed.ncbi.nlm.nih.gov/2270237/> (Accessed: 20 February 2022).

Lakoff, G. and Johnson, M. (2008) *Metaphors we live by*. University of Chicago press.

Lee, D. (2016) *Tay: Microsoft issues apology over racist chatbot fiasco* - BBC News. Available at: <https://www.bbc.com/news/technology-35902104> (Accessed: 23 December 2021).

Lenglet, M. (2011) 'Conflicting Codes and Codings How Algorithmic Trading Is Reshaping Financial Regulation', *Theory, Culture and Society*, 28, pp. 44–66. doi:10.1177/0263276411417444.

Levesque, H.J. (2017) *Common Sense, the Turing Test, and the Quest for Real AI*. MIT Press.

Lianos, M. and with Mary Douglas (2000) 'Dangerization and the End of Deviance: The Institutional Environment', *The British Journal of Criminology*, 40(2), pp. 261–278. doi:10.1093/bjc/40.2.261.

Liu, Z. (2021) 'Sociological perspectives on artificial intelligence: A typological reading', *Sociology Compass*, 15(3), p. e12851. doi:10.1111/soc4.12851.

Loi, D. et al. (2019) 'Co-designing AI futures: Integrating AI ethics, social computing, and design', in *Companion publication of the 2019 on designing interactive systems conference 2019 companion*, pp. 381–384.

Long, D. and Magerko, B. (2020) 'What is AI literacy? Competencies and design considerations', in *Proceedings of the 2020 CHI conference on human factors in computing systems*, pp. 1–16.

Madakam, S., Ramaswamy, R. and Tripathi, S. (2015) 'Internet of Things (IoT): A Literature Review', *Journal of Computer and Communications*, 3, pp. 164–173. doi:10.4236/jcc.2015.35021.

- Marcus, A. (2002) 'The cult of cute: the challenge of user experience design', *interactions*, 9(6). doi:10.1145/581951.581966.
- Marres, N. (2020a) 'Co-existence or displacement: Do street trials of intelligent vehicles test society?', *The British Journal of Sociology*, 71(3), pp. 537–555. doi:10.1111/1468-4446.12730.
- Marres, N. (2020b) 'What If Nothing Happens? Street Trials of Intelligent Cars as Experiments in Participation', in Maasen, S., Dickel, S., and Schneider, C. (eds) *TechnoScienceSociety: Technological Reconfigurations of Science and Society*. Cham: Springer International Publishing (Sociology of the Sciences Yearbook), pp. 111–130. doi:10.1007/978-3-030-43965-1_7.
- Masoudi, N. *et al.* (2019) 'A review of affordances and affordance-based design to address usability', in *Proceedings of the Design Society: International Conference on Engineering Design*. Cambridge University Press, pp. 1353–1362.
- McCarthy, J. *et al.* (1955) 'A Proposal For The Dartmouth Summer Research Project On Artificial Intelligence', p. 13.
- McCarthy, J. (2007) 'What is Artificial Intelligence?', p. 15.
- Mcgreneire, J. and Ho, W. (2000) 'Affordances: Clarifying and Evolving a Concep.', in *Proceedings of the Graphics Interface 2000 Conference*, pp. 179–186.
- Medvedev, A.N., Lambiotte, R. and Delvenne, J.-C. (2019) 'The Anatomy of Reddit: An Overview of Academic Research', in F. Ghanbarnejad *et al.* (eds) *Dynamics On and Of Complex Networks III*. Cham: Springer International Publishing (Springer Proceedings in Complexity), pp. 183–204. Available at: https://doi.org/10.1007/978-3-030-14683-2_9.
- Metin, I. and Camgoz, S.M. (2011) 'The Advances in the History of Cognitive Dissonance Theory', *International Journal of Humanities and Social Science*, 1(6), p. 6.
- Mitchell, R.W. (ed.) (1996) *Anthropomorphism, Anecdotes, and Animals*. Albany: State University of New York Press.

()

- Moore, R.K. (2017) 'Appropriate Voices for Artefacts: Some Key Insights', p. 5.
- Mori, M., MacDorman, K.F. and Kageki, N. (2012) 'The Uncanny Valley [From the Field]', *IEEE Robotics Automation Magazine*, 19(2), pp. 98–100. doi:10.1109/MRA.2012.2192811.
- Mubin, O. and Bartneck, C. (2015) 'Do as I say: exploring human response to a predictable and unpredictable robot', in *Proceedings of the 2015 British HCI Conference on - British HCI '15. the 2015 British HCI Conference*, Lincoln, Lincolnshire, United Kingdom: ACM Press, pp. 110–116. doi:10.1145/2783446.2783582.
- Nadkarni, P.M., Ohno-Machado, L. and Chapman, W.W. (2011) 'Natural language processing: an introduction', *Journal of the American Medical Informatics Association*, 18(5), pp. 544–551. Available at: <https://doi.org/10.1136/amiajnl-2011-000464>.
- Nass, C. and Moon, Y. (2000) 'Machines and Mindlessness: Social Responses to Computers', *Journal of Social Issues*, 56(1), pp. 81–103. doi:10.1111/0022-4537.00153.
- Nass, C., Moon, Y. and Green, N. (1997) 'Are machines gender neutral? Gender-stereotypic responses to computers with voices', *Journal of Applied Social Psychology*, 27(10), pp. 864–876. doi:10.1111/j.1559-1816.1997.tb00275.x.
- Nickerson, R.S. (1999) 'How we know—and sometimes misjudge—what others know: Imputing one's own knowledge to others', *Psychological Bulletin*, 125(6), pp. 737–759. doi:10.1037/0033-2909.125.6.737.
- Norman, D. (1988) *The Psychology Of Everyday Things*. 1st edition. New York: Basic Books.
- Odom, W., Sellen, A., et al. (2012) 'Lost in translation: understanding the possession of digital things in the cloud', in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12. the 2012 ACM annual conference*, Austin, Texas, USA: ACM Press, p. 781. doi:10.1145/2207676.2207789.
- Odom, W., Banks, R., et al. (2012) 'Technology heirlooms?: considerations for

- passing down and inheriting digital materials', in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12. the 2012 ACM annual conference*, Austin, Texas, USA: ACM Press, p. 337. doi:10.1145/2207676.2207723.
- Patterson, M. (1983) *Nonverbal Behavior: A Functional Perspective*. New York: Springer-Verlag (Springer Series in Social Psychology). doi:10.1007/978-1-4612-5564-2.
- Phan, T. (2019) 'Amazon Echo and the Aesthetics of Whiteness', *Catalyst: Feminism, Theory, Technoscience*, 5(1), pp. 1–38. doi:10.28968/cftt.v5i1.29586.
- Pierce, J. and DiSalvo, C. (2018) 'Addressing Network Anxieties with Alternative Design Metaphors', in *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '18), pp. 1–13. Available at: <https://doi.org/10.1145/3173574.3174123>.
- Posavec, S. and McInerney, G. (2009) *The Evolution of The Origin of Species, Visual Complexity*. Available at: <http://www.visualcomplexity.com/vc/project.cfm?id=696> (Accessed: 20 February 2022).
- Praat (2022) *Softonic*. Available at: <https://praat.en.softonic.com/mac> (Accessed: 19 November 2022).
- Proudfoot, D. (2011) 'Anthropomorphism and AI: Turing's much misunderstood imitation game', *Artificial Intelligence*, 175(5–6), pp. 950–957. doi:10.1016/j.artint.2011.01.006.
- Read, S.J. and Simon, D. (2012) 'Parallel constraint satisfaction as a mechanism for cognitive consistency', in *Cognitive consistency: A fundamental principle in social cognition*. New York, NY, US: Guilford Press, pp. 66–86.
- Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge University Press.
- Riek, L.D. et al. (2009) 'How anthropomorphism affects empathy toward robots',

()

in *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, pp. 245–246.

Rohrer, R., Sibert, J. and Ebert, D. (1998) 'The Shape of Shakespeare: Visualizing Text using Implicit Surfaces', in, pp. 121–129. doi:10.1109/INFVIS.1998.729568.

Rowland, C. *et al.* (2015) *Designing Connected Products: UX for the Consumer Internet of Things*. 1 edition. Sebastopol: O'Reilly Media.

Rundle, J. (2019) 'Wall Street Braces for Impact of AI', *Wall Street Journal*, 9 December. Available at: <https://www.wsj.com/articles/wall-street-braces-for-impact-of-ai-11575887402> (Accessed: 23 December 2021).

Ryan, E.B. and Cacioppo, T. (1981) 'Dimensions of Sociolinguistics - K. R. Scherer and H. Giles (eds.), Social markers in speech. Cambridge University Press, 1979. Pp. 395.', *Language in Society*, 10(3), pp. 443–448. doi:10.1017/S0047404500008873.

Šabanović, S. (2014) 'Inventing Japan's "robotics culture": The repeated assembly of science, technology, and culture in social robotics', *Social Studies of Science*, 44(3), pp. 342–367. doi:10.1177/0306312713509704.

Sau-lai Lee *et al.* (2005) 'Human Mental Models of Humanoid Robots', in *Proceedings of the 2005 IEEE International Conference on Robotics and Automation. 2005 IEEE International Conference on Robotics and Automation*, Barcelona, Spain: IEEE, pp. 2767–2772. doi:10.1109/ROBOT.2005.1570532.

Schwartz, R.D. (1989) 'Artificial Intelligence as a Sociological Phenomenon', *The Canadian Journal of Sociology / Cahiers canadiens de sociologie*, 14(2), pp. 179–202. doi:10.2307/3341290.

Shim, M., Johnson, B. and Gasson, S. (2020) *A Mixed Methods–Grounded Theory Design for Producing More Refined Theoretical Models*. Available at: <https://journals.sagepub.com/doi/full/10.1177/1558689820932311> (Accessed: 19 November 2022).

Shneiderman, B. (1989) *A nonanthropomorphic style guide: overcoming*

the humpty dumpty syndrome. Available at: <https://www.semantic-scholar.org/paper/A-nonanthropomorphic-style-guide%3A-overcoming-the-Shneiderman/Ofaafa651e34989c9ffe0b72560b162eecd693> (Accessed: 31 December 2021).

Skeldon, P. (2018) *Amazon Echo dominates UK smart speaker market, but all are growing*, *InternetRetailing*. Available at: <https://internetretailing.net/mobile-theme/mobile-theme/amazon-echo-dominates-uk-smart-speaker-market-but-all-are-growing-18873> (Accessed: 31 December 2021).

Sloane, M. and Moss, E. (2019) 'AI's social sciences deficit', *Nature Machine Intelligence*, 1(8), pp. 330–331. doi:10.1038/s42256-019-0084-6.

Spencer, D. (2016) *Work in and beyond the Second Machine Age: the politics of production and digital technologies - David Spencer, 2017*. Available at: <https://journals.sagepub.com/doi/10.1177/0950017016645716> (Accessed: 23 December 2021).

Spigel, L. (2001) *Welcome to the Dreamhouse: Popular Media and Postwar Suburbs*. Illustrated edition. Durham, NC: Duke University Press Books.

Sproull, L. et al. (1996) 'When the Interface Is a Face', *Human-Computer Interaction*, 11(2), pp. 97–124. doi:10.1207/s15327051hci1102_1.

SPSS Software | IBM (2019). Available at: <https://www.ibm.com/analytics/spss-statistics-software> (Accessed: 23 September 2019).

Steele, C. (1988) *The Psychology of Self-Affirmation: Sustaining the Integrity of the Self - ScienceDirect*. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0065260108602294> (Accessed: 20 February 2022).

Stilgoe, J. (2018) 'Machine learning, social learning and the governance of self-driving cars', *Social Studies of Science*, 48(1), pp. 25–56. doi:10.1177/0306312717741687.

Suchman, L. (2006) *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd edition. Cambridge ; New York: Cambridge University Press.

()

Suchman, L. (2011) 'Subject objects', *Feminist Theory*, 12, pp. 119–145. doi:10.1177/1464700111404205.

Suchman, L.A. (1987) *Plans and Situated Actions: The Problem of Human-Machine Communication*. 2nd edition. Cambridge Cambridgeshire ; New York: Cambridge University Press.

SurveyMonkey: The World's Most Popular Free Online Survey Tool (2018). Available at: <https://www.surveymonkey.com/> (Accessed: 23 September 2019).

Syrdal, D.S. et al. (2008) 'Sharing Spaces with Robots in a Home Scenario - Anthropomorphic Attributions and their Effect on Proxemic Expectations and Evaluations in a Live HRI Trial', *AAAI Fall Symposium - Technical Report* [Preprint].

Sztandar-Sztanderska, K. and Zielenska, M. (2018) 'Changing social citizenship through information technology', *Social Work & Society*, 16.

Sztandar-Sztanderska, K. and Zielenska, M. (2020) 'What Makes an Ideal Unemployed Person? Values and Norms Encapsulated in a Computerized Profiling Tool 1', *Social Work & Society*, 18.

Tamir, P. and Zohar, A. (1991) 'Anthropomorphism and teleology in reasoning about biological phenomena', *Science Education*, 75(1), pp. 57–67.

Taplin, J. (2017) *Move Fast and Break Things: How Facebook, Google, and Amazon Cornered Culture and Undermined Democracy*. New York: Little Brown and Company.

Taylor, A.S. (2009) 'Machine intelligence', in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: Association for Computing Machinery (CHI '09), pp. 2109–2118. Available at: <https://doi.org/10.1145/1518701.1519022>.

Topol, E.J. (2019) 'High-performance medicine: the convergence of human and artificial intelligence', *Nature Medicine*, 25(1), pp. 44–56. doi:10.1038/s41591-018-0300-7.

Turkle, S. (2005) *The Second Self: Computers and the Human Spirit*. Twentieth Anniversary edition. Cambridge, Mass: MIT Press.

UNESCO Digital Library (2019) *I'd blush if I could: closing gender divides in digital skills through education* -. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000367416.page=1> (Accessed: 13 February 2020).

Urquhart, C. (2003) 'Regrounding Grounded Theory-or Reinforcing Old Prejudices? A Brief Reply to Bryant', *JITTA* [Preprint]. Available at: https://www.academia.edu/400508/Regrounding_Grounded_Theory_or_Reinforcing_Old_Prejudices_A_Brief_Reply_to_Bryant (Accessed: 24 November 2021).

Urquiza-Haas, E.G. and Kotrschal, K. (2015) 'The mind behind anthropomorphic thinking: attribution of mental states to other species', *Animal Behaviour*, 109, pp. 167–176. doi:10.1016/j.anbehav.2015.08.011.

Viégas, F. (1999) *Chat Circles*. Available at: <http://alumni.media.mit.edu/~fviegas/projects/chatcircles/index.htm> (Accessed: 8 January 2022).

de Visser, E.J. *et al.* (2016) 'Almost human: Anthropomorphism increases trust resilience in cognitive agents', *Journal of Experimental Psychology: Applied*, 22(3), pp. 331–349. doi:10.1037/xap0000092.

Walker, J.H., Sproull, L. and Subramani, R. (1994) 'Using a human face in an interface', in *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94. the SIGCHI conference*, Boston, Massachusetts, United States: ACM Press, pp. 85–91. doi:10.1145/191666.191708.

Wasserman, T. (2014) *Amazon Echo is another step toward commoditizing your personal life*, *Mashable*. Available at: <https://mashable.com/archive/amazon-echo-commoditizing> (Accessed: 31 December 2021).

Watson, D. (2019) 'The rhetoric and reality of anthropomorphism in artificial intelligence', *Minds and Machines*, 29(3), pp. 417–440.

Watson, D.S. *et al.* (2019) 'Clinical applications of machine learning algorithms: beyond the black box', *BMJ (Clinical research ed.)*, 364, p. 1886. doi:10.1136/

()

bmj.l886.

Waytz, A. et al. (2010) 'Making sense by making sentient: Effectance motivation increases anthropomorphism.', *Journal of Personality and Social Psychology*, 99(3), pp. 410–435. doi:10.1037/a0020240.

Waytz, A., Cacioppo, J. and Epley, N. (2010) 'Who sees human? The stability and importance of individual differences in anthropomorphism', *Perspectives on Psychological Science*, 5(3), pp. 219–232.

Waytz, A., Epley, N. and Cacioppo, J.T. (2010) 'Social Cognition Unbound: Insights Into Anthropomorphism and Dehumanization', *Current Directions in Psychological Science*, 19(1), pp. 58–62. doi:10.1177/0963721409359302.

Waytz, A., Heafner, J. and Epley, N. (2014) 'The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle', *Journal of Experimental Social Psychology*, 52, pp. 113–117. doi:10.1016/j.jesp.2014.01.005.

Weiser, M. (1991) 'The Computer for the 21st Century', *Scientific American*, pp. 66–75.

White, R.W. (1959) 'Motivation reconsidered: the concept of competence.', *Psychological review*, 66, pp. 297–333. doi:10.1037/h0040934.

Wiafe, I., Nakata, K. and Gulliver, S. (2014) 'Categorizing users in behavior change support systems based on cognitive dissonance', *Personal and ubiquitous computing*, 18(7), pp. 1677–1687.

Wicklund, R.A. and Brehm, J.W. (2013) *Perspectives on cognitive dissonance*. Psychology Press.

Wilson, B., Hoffman, J. and Morgenstern, J. (2019) 'Predictive Inequity in Object Detection', *arXiv:1902.11097 [cs, stat]* [Preprint]. Available at: <http://arxiv.org/abs/1902.11097> (Accessed: 23 December 2021).

Wilson, S.G. and Haslam, N. (2013) 'Humanness beliefs about behavior: An index and comparative human-nonhuman behavior judgments', *Behavior Research Methods*, 45(2), pp. 372–382. doi:10.3758/s13428-012-0252-7.

- Wise, J. et al. (1995) 'Visualizing the non-Visual: Spatial Analysis and Interaction With Information from Text Documents', in: *Proceedings of the Information Visualization Conference*, pp. 51–58. doi:10.1109/INFVIS.1995.528686.
- Wolf, T. et al. (2020) 'Transformers: State-of-the-Art Natural Language Processing', in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Online: Association for Computational Linguistics, pp. 38–45. Available at: <https://doi.org/10.18653/v1/2020.emnlp-demos.6>.
- Wong, R. (2015) *Amazon Echo: This is what a smart home should feel like*. Available at: <https://mashable.com/archive/amazon-echo-review> (Accessed: 31 December 2021).
- Xu, W. (2019) 'Toward human-centered AI: a perspective from human-computer interaction', *interactions*, 26(4), pp. 42–46.
- You, S. and Robert, L. (2018) 'Teaming up with robots: An IMOI (inputs-mediators-outputs-inputs) framework of human-robot teamwork', *You, S. and Robert, LP (2017). Teaming Up with Robots: An IMOI (Inputs-Mediators-Outputs-Inputs) Framework of Human-Robot Teamwork, International Journal of Robotic Engineering, (IJRE), 2(3)*.
- Zhang, K. and Aslan, A.B. (2021) 'AI technologies for education: Recent research & future directions', *Computers and Education: Artificial Intelligence*, 2, p. 100025. doi:10.1016/j.caeai.2021.100025.
- Złotowski, J. et al. (2015) 'Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction', *International Journal of Social Robotics*, 7(3), pp. 347–360. doi:10.1007/s12369-014-0267-6.
- Złotowski, J., Strasser, E. and Bartneck, C. (2014) 'Dimensions of anthropomorphism: from humanness to humanlikeness', in *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14. the 2014 ACM/IEEE international conference*, Bielefeld, Germany: ACM Press, pp. 66–73. doi:10.1145/2559636.2559679.

05 APPENDICES

A. GLOSSARY OF TERMS

The glossary of terms has been generated with GPT-3, and curated, where appropriate, to match the use of terms in this work. Source code is included in Appendix O.

Affordance: An object or environment's ability to provide a cue for an action.

Algorithm: A step-by-step (here, computational) procedure for solving a problem or accomplishing a task. It is a specific set of instructions to be followed when solving a problem.

Anthropomorphism: the (here cognitive) attribution of human features or characteristics to nonhuman entities.

Automation: In computing, automation is the use of software to control the execution of tasks normally performed by a human operator. Automation software can automate small or repetitive tasks, freeing up the operator to perform more complex or high-value tasks.

Autonomy: The independence and freedom of a person/group/agent to act on their own behalf.

Blackbox: A device or system whose workings are not immediately apparent to the user, and which is either difficult or impossible to understand or analyse.

Chatbot: A computer program that can mimic human conversation.

Cognitive dissonance: The mental discomfort experienced when simultaneously holding two or more conflicting cognitions, i.e., when one holds two or more contradictory beliefs, ideas, or values.

Computer vision: A field of artificial intelligence involving the development of methods and algorithms that allow computers to "see" and interpret the surrounding world.

Conceptual model: An intentional model of a system or concept that shows the relationships between the elements of the system.

Conceptualisation: The process of designing a representation of something. This might be a physical object, like a chair, or something more abstract, like a concept, idea or system.

Conversational Analysis: A research methodology that looks at the structure of spoken and written language in order to understand the social interaction that is taking place.

Dataset: A collection of data that is organised for use in a computer system.

Deep Learning: A subset of machine learning that is inspired by the workings of the human brain. It involves the use of artificial neural networks, which are systems that can learn to recognize patterns of data, in order to make predictions or decisions. Deep learning networks can be extremely effective in tasks such as image recognition, natural language processing, and predictive modelling.

Effectance Motivation: A desire to maintain or improve one's position in the environment, characterized by a need to feel competent and effective.

Elicit Agent Knowledge: Refers to the acquisition of knowledge about a particular agent. This knowledge can be used to inform decision-making processes or to improve interactions with the agents in question.

Experimental Research: A type of research in which the researcher manipulates the independent variable and observes the effect on the dependent variable.

Grounded Theory: A research methodology that aims to inductively develop a theory (usually of a social phenomenon) grounded in data.

IoT: An abbreviation for "Internet of Things." The phrase refers to the growing network of physical objects and devices — from cars and watches to home appliances and industrial machinery — that are connected to the internet and can communicate with one another.

Machine Learning: A subset of artificial intelligence that enables computers to learn from data, without being explicitly programmed to do so.

Mental model: An explanatory framework for understanding how something works. It is a simplified representation of reality, constructed in our minds, that helps us make sense of the environment.

Natural Language Processing: A subfield of computer science and artificial intelligence involving the study of natural language and its various computational aspects. NLP research focuses on understanding natural language communication with computers, extracting meaning from text, and generating natural language text.

Perception: The process of organising and interpreting sensory information in order to understand the environment.

Personification: When an object or nonhuman entity is given human qualities.

Schema: A cognitive structure that organises knowledge and experiences. Synonymous with a mental model.

Signified: A signified is the meaning or concept that a signifier represents.

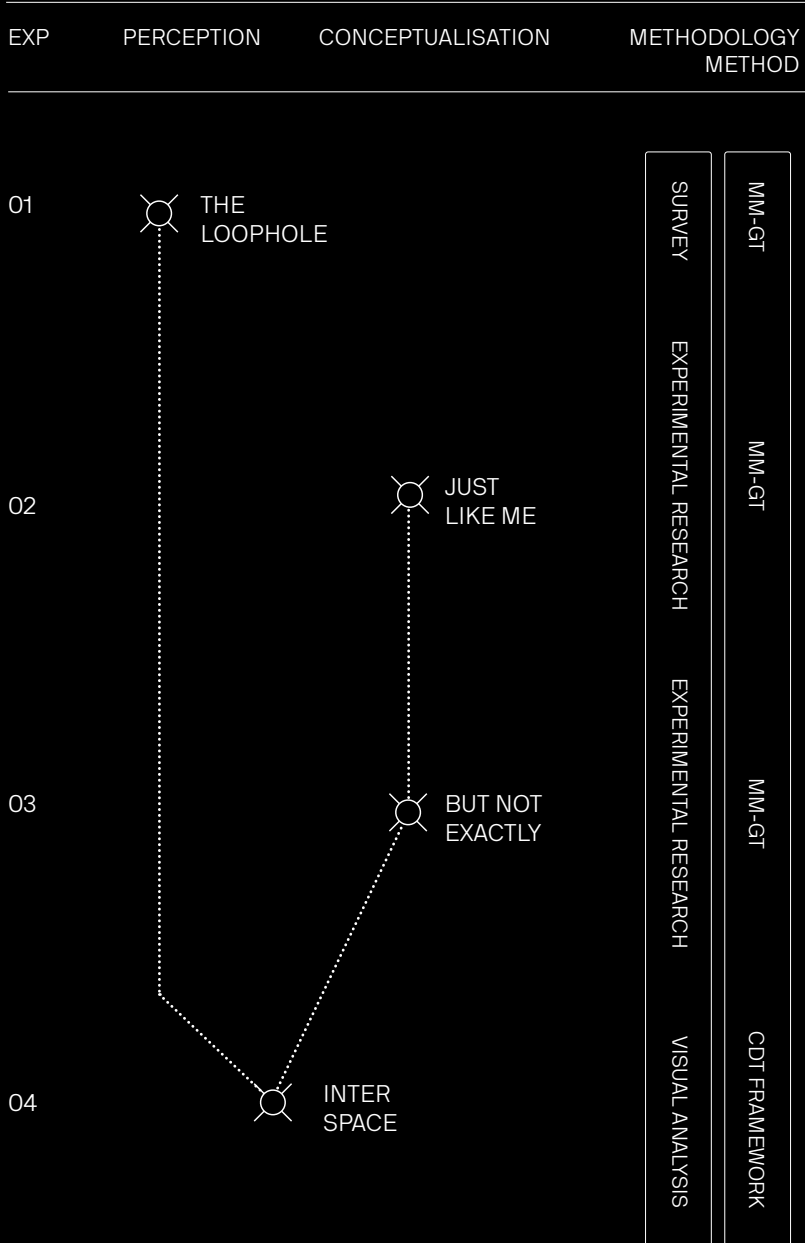
Signifier: In semiotics, a signifier is anything that stands for, or represents, something else. In other words, it is the form that a sign takes. The word “signifier” is derived from the Latin word “signum,” which means “sign.”

Sociality: The quality or state of being social or the degree of social interaction.

Suspension of disbelief: A state of mind in which a person accepts as true the unbelievable or the impossible.

Symbolic Interactionism: A theoretical perspective that views society as the product of the everyday interactions of individuals and the meaning they attach to their social world. In other words, it is the study of how humans interact with each other through the use of symbols (including language, gestures, and objects).

B. PHD TIMELINE



C. VIRTUAL ASSISTANTS SURVEY

SECTION 1

Q1. Please indicate your gender (optional):

- Female Male Non-Binary Prefer not to say

Q2. What kind of virtual assistant do you have? (e.g., Amazon Echo, Google Home, Invoke, Sonos One etc.)

Q3. How long have you had it for?

- Less than a month
 1-3 months
 3-6 months
 6 months to a year
 More than a year

Q4. How often do you use it?

- Very often (multiple times a day)
 Often (almost every day)
 Moderately (a couple of times a week)
 Rarely (when I remember)
 Never

Q5. How often do you use the following features?

Play Music:

- Always Often Sometimes Rarely Never

News:

- Always Often Sometimes Rarely Never

Control Smart Home Devices:

- Always Often Sometimes Rarely Never

Weather or Traffic:

- Always Often Sometimes Rarely Never

Answer General Knowledge Questions:

- Always Often Sometimes Rarely Never

Order Online:

- Always Often Sometimes Rarely Never

Chat (e.g.: tell jokes):

- Always Often Sometimes Rarely Never

Time management (alarm, timer, schedule):

- Always Often Sometimes Rarely Never

Other (Please Specify):

SECTION 2**Q6. How much do you agree with the following statements?****My Virtual Assistant meets my needs exactly as I expected:**

- Completely Agree
 Mostly Agree
 Neither Agree Nor Disagree
 Mostly Agree
 Completely Disagree

My Virtual Assistant has a robust design

- Completely Agree
 Mostly Agree
 Neither Agree Nor Disagree
 Mostly Agree
 Completely Disagree

My Virtual Assistant is responsive to all my questions and requests

- Completely Agree
- Mostly Agree
- Neither Agree Nor Disagree
- Mostly Agree
- Completely Disagree

Q7. How often do you ___ when interacting with your virtual assistant?**Use polite words (please, thank you, sorry etc.)**

- Always Often Sometimes Rarely Never

Speak in a natural way (as with a friend)

- Always Often Sometimes Rarely Never

Give abbreviated commands (e.g. "play music")

- Always Often Sometimes Rarely Never

Raise your voice

- Always Often Sometimes Rarely Never

Use swear words

- Always Often Sometimes Rarely Never

Use aggressive behaviour (e.g.: switch it off, put it away etc.)

- Always Often Sometimes Rarely Never

Q8. How well do you think you understand the technology behind it?

(1 = Not at all; 10 = Extremely proficient)

- 1 2 3 4 5 6 7 8 9 10

SECTION 3

Q9. How applicable are the following statements? My virtual assistant is ___
(1 = NA; 5 = Very Applicable)

Polite

1 2 3 4 5

Sophisticated

1 2 3 4 5

Moral

1 2 3 4 5

Rational

1 2 3 4 5

Mature

1 2 3 4 5

Empathetic (emotionally responsive)

1 2 3 4 5

Friendly

1 2 3 4 5

Curious

1 2 3 4 5

Compliant / Flexible

1 2 3 4 5

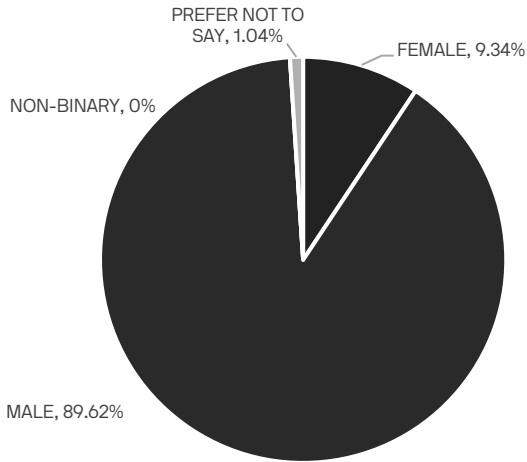
Autonomous

1 2 3 4 5

Q10. How would you describe your virtual assistant in a couple of words / sentences?

D. VA SURVEY SUMMARY

Q1. Please indicate your gender (optional):



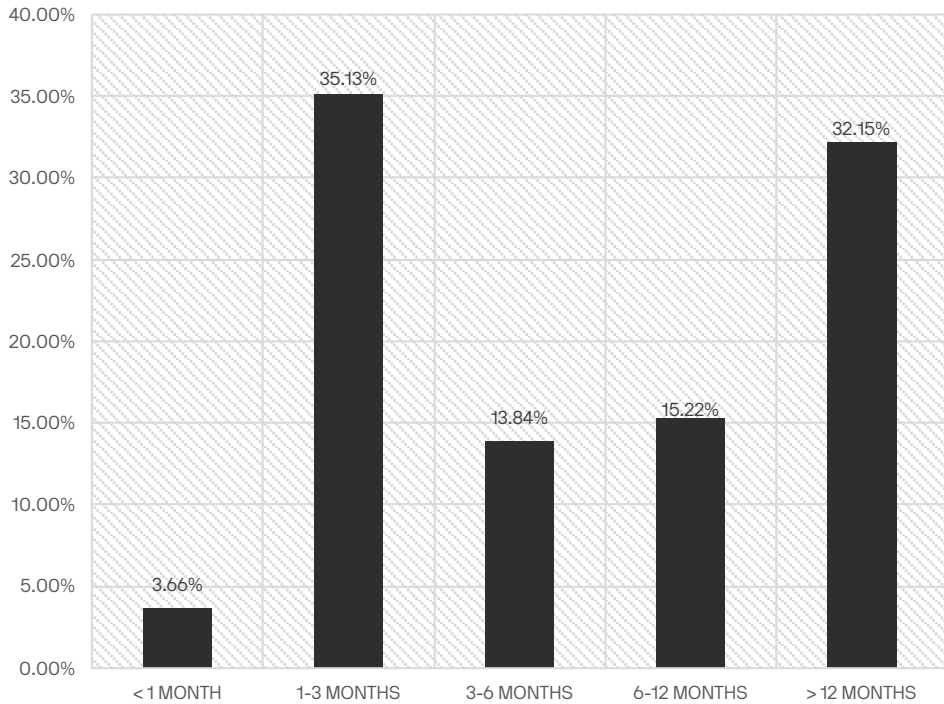
Answered: 867 Skipped: 7

Answer Choices	Responses	
Female (1)	9.34%	81
Male (2)	89.62%	777
Non-Binary (3)	0%	0
Prefer not to say (4)	1.04%	9

Basic Statistics

Minimum	Maximum	Median	Mean	Standard Deviation
1.00	4.00	2.00	1.92	0.31

Q3. How long have you had it for?



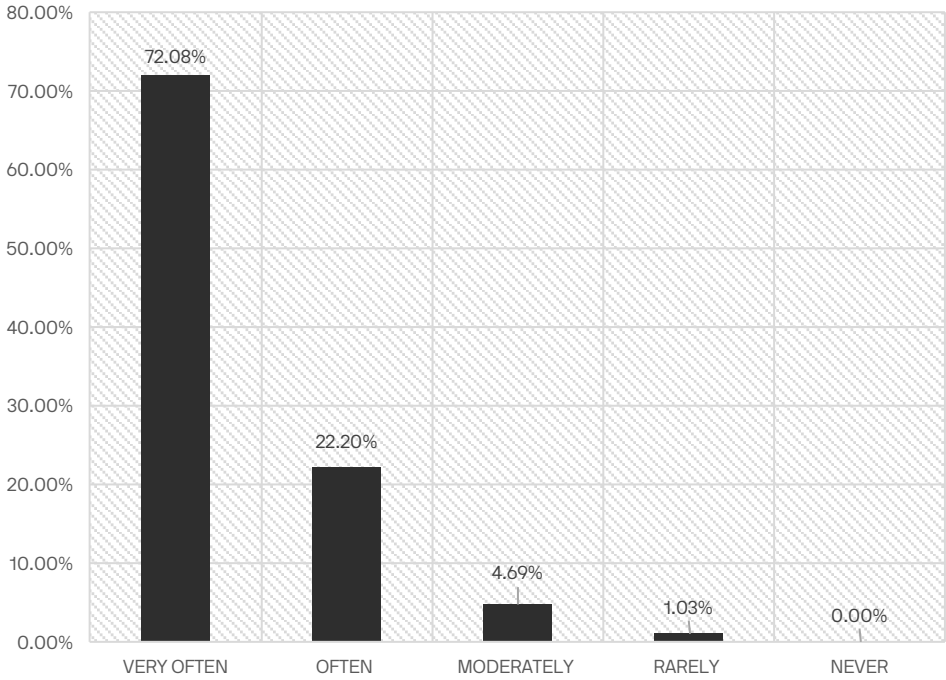
Answered: 874 Skipped: 0

Answer Choices	Responses	
Less than a month (1)	3.66%	32
1-3 months (2)	35.13%	307
3-6 months (3)	13.84%	121
6 months to a year (4)	15.22%	133
More than a year (5)	32.15%	281

Basic Statistics

Minimum	Maximum	Median	Mean	Standard Deviation
1.00	5.00	3.00	3.37	1.34

Q4. How often do you use it?



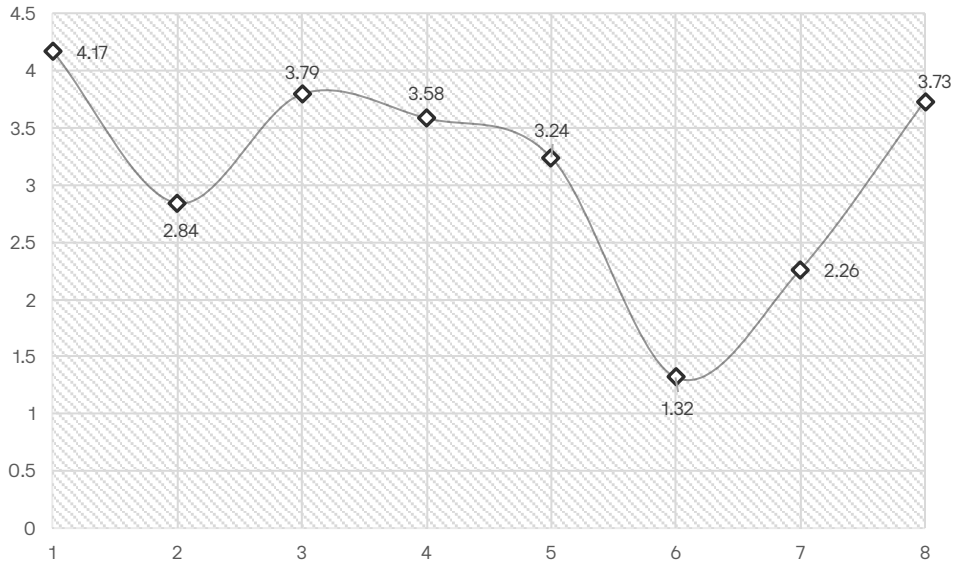
Answered: 874 Skipped: 0

Answer Choices	Responses	
Very often (multiple times a day) (1)	72.08%	630
Often (almost every day) (2)	22.20%	194
Moderately (a couple of times a week) (3)	4.69%	41
Rarely (when I remember) (4)	1.03%	9
Never (5)	0.00%	0

Basic Statistics

Minimum	Maximum	Median	Mean	Standard Deviation
1.00	4.00	1.00	1.35	0.62

Q5. How often do you use the following features?



Answered: 874 Skipped: 0

	Never	Rarely	Sometimes	Often	Always	Weighted Average
Play Music (1)	1.26% 11	3.20% 28	12.59% 110	43.48% 380	39.47% 345	4.17
News (2)	12.47% 109	26.09% 228	31.01% 271	25.86% 226	4.58% 40	2.84
Control Smart Home (3)	15.79% 138	4.23% 37	9.15% 80	26.66% 233	44.16% 386	3.79
Weather Traffic (4)	2.86% 25	10.64% 93	23.57% 206	51.03% 446	11.90% 104	3.58
Answer Questions (5)	2.40% 21	12.81% 112	45.77% 400	36.04% 315	2.97% 26	3.24
Oder Online (6)	75.51% 660	17.85% 156	5.61% 49	1.03% 9	0.00% 0	1.32
Chat (7)	19.34% 169	45.65% 399	25.74% 225	8.24% 72	1.03% 9	2.26
Alarm, Timer etc. (8)	2.29% 20	6.75% 59	23.80% 208	49.54% 433	17.62% 154	3.73

Basic Statistics

	Minimum	Maximum	Median	Mean	Standard Deviation
Play Music (1)	1.00	5.00	4.00	4.17	0.86
News (2)	1.00	5.00	3.00	2.84	1.08
Control Smart Home (3)	1.00	5.00	4.00	3.79	1.44
Weather Traffic (4)	1.00	5.00	4.00	3.58	0.93
Answer Questions (5)	1.00	5.00	3.00	3.24	0.80
Oder Online (6)	1.00	5.00	1.00	1.32	0.63
Chat (7)	1.00	5.00	2.00	2.26	0.90
Alarm, Timer etc. (8)	1.00	5.00	4.00	3.73	0.91

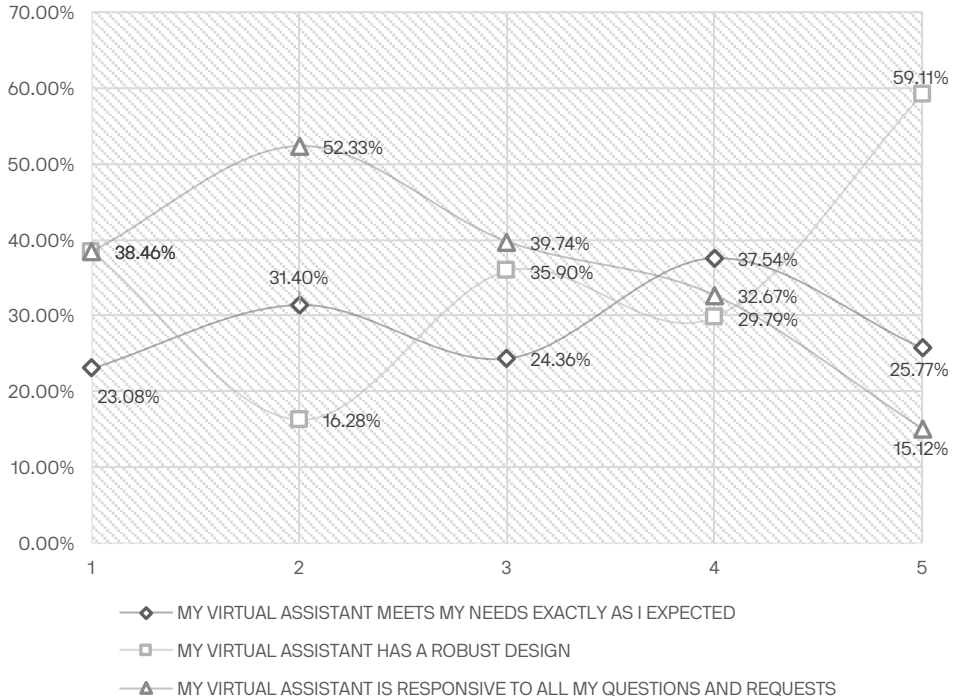
#	Other (Please Specify)	Date
1	Communicate with family in another country	2018-02-07 17:08:19
2	Set Reminded and Calendar appointments	2018-02-07 12:39:02
3	Podcasts	2018-02-07 10:24:59
4	Heard people listen to audiobooks- might try that.	2018-02-07 10:24:33
5	Games and learning French	2018-02-07 10:05:11
6	Keeping shopping lists and other reminders - Often; Recipes for cooking - Sometimes	2018-02-07 9:40:33
7	Add items to shopping list and to-do list	2018-02-07 9:17:07
8	Sports Update	2018-02-07 7:42:09
9	I developer Alexa skills so a lot of testing	2018-02-07 6:39:38
10	Ordering Online is unavailable in my country or I would likely use it "sometimes"	2018-02-07 5:41:03
11	Games and trivia, sometimes	2018-02-07 5:33:27
12	Find the phone (this is the main reason of purchase seriously)	2018-02-07 5:27:32
13	audible	2018-02-07 5:08:39
14	Time wasting by playing trivia games or asking random questions.	2018-02-07 4:57:06
15	Ambient Noise	2018-02-07 4:39:29
16	"Good morning" gets a daily trivia	2018-02-07 3:44:23

17	Listen to Podcasts	2018-02-07 3:52:41
18	Audio books and games (sometimes)	2018-02-07 3:32:28
19	Alexa Dot uses Bing, so it is mostly useless for any searches or intelligence.	2018-02-07 3:31:26
20	Use it to learn weather, turn on lights, play music	2018-02-07 3:19:23
21	Play ambient noises for sleep & relaxation	2018-02-07 3:07:44
22	Grocery list (often)	2018-02-07 3:04:18
23	I use the broadcast feature many times a day to talk throughout the house.	2018-02-07 3:03:31
24	White noise/relaxation sounds before bedtime	2018-02-07 2:51:30
25	Development/hacks	2018-02-07 2:35:30
26	Audio Books	2018-02-07 2:17:32
27	Control FireTV Stick video playback	2018-02-07 2:12:23
28	Ambient noise, broadcast as intercom	2018-02-07 2:08:37
29	drop in	2018-02-07 2:05:37
30	Netflix / YouTube	2018-02-07 1:59:33
31	I also do games, like the Magic Door.	2018-02-07 1:55:18
32	Make a phone call (rarely).	2018-02-07 1:54:36
33	Have hues lights and nest and other things linked to it	2018-02-07 1:16:54
34	Play ambient noise - this is it's main purpose	2018-02-07 1:08:19
35	broadcast to other rooms in home (from house or away from home)	2018-02-07 0:58:45
36	Check BMW status	2018-02-07 0:50:15
37	Note: I'm from the EU, so cant use it fully yet.	2018-02-07 0:31:47
38	Rarely games, like Jeopardy.	2018-02-07 0:19:20
39	I live in UK so don't think I can online order or use as phone yet	2018-02-07 0:05:08
40	Digital Radio	2018-02-06 23:21:15
41	Controlling Tesla.	2018-02-06 22:54:14
42	playing white noise while I sleep	2018-02-06 23:56:59
43	Games	2018-02-06 22:37:23

44	Call/Text other people	2018-02-06 22:33:32
45	Play Games, In-home intercom,	2018-02-06 22:21:43
46	Unable to order online in UK but would use often.	2018-02-06 22:15:40
47	Broadcast, I use that sometimes	2018-02-06 22:09:42
48	I try to find random quirks that the assistant has	2018-02-06 21:50:38
49	It's essentially a voice remote control + timers. Neither asking something nor doing shopping is not appealing to me without screen (I need context)	2018-02-06 21:27:33
50	Home automation and a quick news/weather in the morning.	2018-02-06 21:23:48
51	Personal automation outside of smart home devices like IFTTT and Tasker for Android	2018-02-06 20:24:29
52	Grocery list	2018-02-06 20:21:58
53	Podcasts	2018-02-06 20:19:53
54	Kind of time management but as timer for our 3 year old to help him switch tasks	2018-02-06 20:21:26
55	Alexa. fart. Alexa bark	2018-02-06 20:16:06
56	Unit conversions and recipes in the kitchen	2018-02-06 20:09:27
57	Play radio specifically, e.g. BBC Radio 2 & 6, Classic FM	2018-02-06 20:00:00
58	Call and/or video chat	2018-02-06 19:43:41
59	make phone calls, Check movie times, stock prices,	2018-02-06 19:34:50
60	Answer Local Questions ("When is the next train to downtown?" "When does the corner grocer close?")	2018-02-06 19:29:31
61	Recipes and covering units when cooking (degrees f to c, cups to grams etc)	2018-02-06 19:21:57
62	Sometimes I use the games while entertaining company.	2018-02-06 19:09:03
63	Pull my finger fart joke	2018-02-06 18:59:11
64	Play podcasts	2018-02-06 18:51:41
65	Audiobooks	2018-02-06 18:38:01
66	Grocery List - Often	2018-02-06 18:28:00
67	I also have Alexa but for this Q/A I'm going to answer with Google Home since I much prefer it	2018-02-06 18:30:02
68	Casting Netflix to TV	2018-02-06 18:19:30

69	I have not purchased any controllers as of yet but I am planning on it soon.	2018-02-06 18:07:13
70	White noise machine	2018-02-06 18:03:49
71	kids games: often	2018-02-06 17:50:17
72	Send messages in the home through the broadcast feature.	2018-02-06 17:50:09
73	Order online isn't fully implemented in Canada to my knowledge	2018-02-06 18:13:21
74	NPR, Rainstorm sounds get used frequently	2018-02-06 17:38:34
75	Control character load outs in games.	2018-02-06 17:36:56
76	Some functions are not available (online shopping) because the device is not officially supported in my country	2018-02-06 17:02:11
77	Podcasts and Phone Calls	2018-02-06 16:54:06
78	"news" insofar as I use it to play NPR, and "music" in that I play audio books	2018-02-06 16:27:13
79	What is my commute to work	2018-02-06 16:29:05
80	Play radio and podcasts	2018-02-06 15:03:25
81	Play audio books	2018-02-06 15:03:57
82	If I owned any smart home appliances I would use it to control these.	2018-02-06 14:32:54
83	etc.: knowing the bin collection schedule (requires add-on component to activate)	2018-02-06 12:58:06

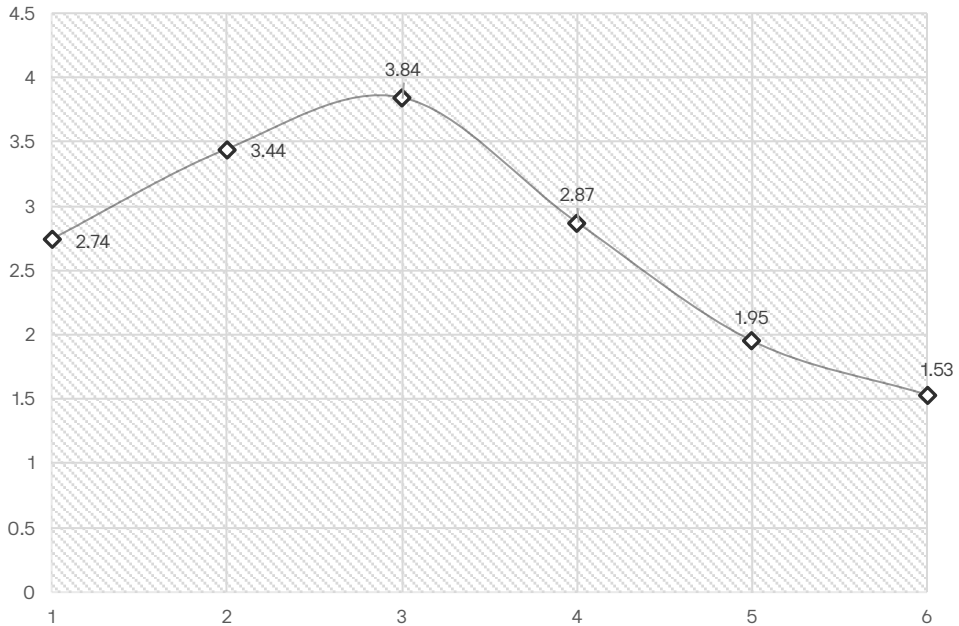
Q6. How much do you agree with the following statements?



Answered: 874 Skipped: 0

	My Virtual Assistant meets my needs exactly as I expected	My Virtual Assistant has a robust design	My Virtual Assistant is responsive to all my questions and requests	Total	Weighted Average
Completely Disagree (1)	23.08% 6	38.46% 10	38.46% 10	26	1.00
Mostly Disagree (2)	31.40% 54	16.28% 28	52.33% 90	172	2.00
Neither agree nor disagree (3)	24.36% 114	35.90% 168	39.74% 186	468	3.00
Mostly Agree (4)	37.54% 625	29.79% 496	32.67% 544	1665	4.00
Completely Agree (5)	25.77% 75	59.11% 172	15.12% 44	291	5.00

Q7. How often do you ___ when interacting with your virtual assistant?



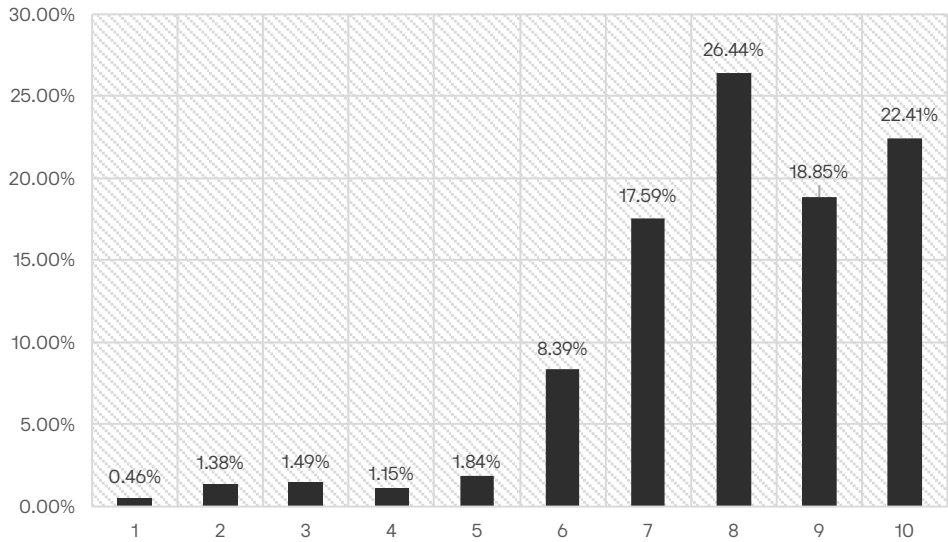
Answered: 874 Skipped: 0

	Never	Rarely	Sometimes	Often	Always	Weighted Average
Use polite words (1)	19.22% 168	23.11% 202	27.92% 244	23.46% 205	6.29% 55	2.74
Speak in a natural way (2)	3.55% 31	13.04% 114	31.81% 278	39.36% 344	12.24% 107	3.44
Give abbreviated commands (3)	3.20% 28	10.64% 93	31.69% 277	43.71% 382	10.76% 94	3.84
Raise your voice (4)	9.50% 83	23.46% 205	40.85% 357	23.34% 204	2.86% 25	2.87
Use swear words (5)	41.08% 359	31.24% 273	19.79% 173	7.32% 64	0.57% 5	1.95
Use aggressive behaviour (6)	65.33% 571	21.28% 186	9.50% 83	3.32% 29	0.57% 5	1.53

Basic Statistics

	Minimum	Maximum	Median	Mean	Standard Deviation
Use polite words (1)	1.00	5.00	3.00	2.74	1.19
Speak in a natural way (2)	1.00	5.00	4.00	3.44	0.98
Give abbreviated commands (3)	1.00	5.00	4.00	3.48	0.93
Raise your voice (4)	1.00	5.00	3.00	2.87	0.97
Use swear words (5)	1.00	5.00	2.00	1.95	0.98
Use aggressive behaviour (6)	1.00	5.00	1.00	1.53	0.84

Q8. How well do you think you understand the technology behind it?
 (1 = Not at all; 10 = Extremely proficient)



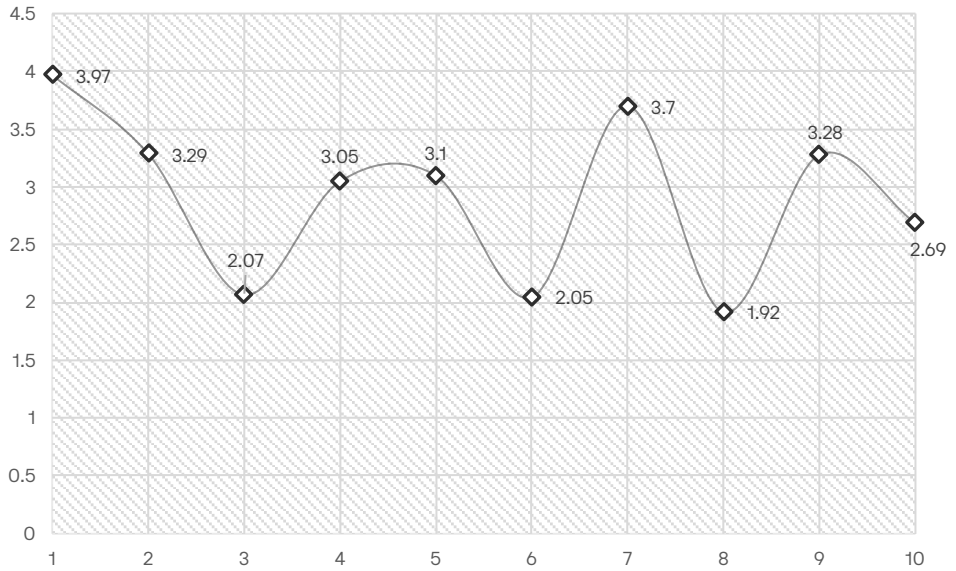
Answered: 874 Skipped: 0

Answer Choices	Responses	
Not at all (1)	0.46%	4
(2)	1.38%	12
(3)	1.49%	13
(4)	1.15%	10
(5)	1.84%	16
(6)	8.39	73
(7)	17.59	153
(8)	26.44	230
(9)	18.85	164
Extremely proficient (10)	22.41	195

Basic Statistics

Minimum	Maximum	Median	Mean	Standard Deviation
1.00	10.00	8.00	7.97	1.82

Q9. How applicable are the following statements? My virtual assistant is ___
(1 = NA; 5 = Very Applicable)



Answered: 874 Skipped: 0

	1 (N/A)	2	3	4	5 (very)	Weighted Average
Polite (1)	5.26% 46	3.43% 30	16.02% 140	39.13% 342	36.16% 316	3.97
Sophisticated (2)	8.24% 72	11.56% 101	33.98% 297	35.70% 312	10.53% 92	3.29
Moral (3)	47.14% 412	15.56% 136	24.49% 214	9.04% 79	3.78% 33	2.07
Rational (4)	16.36% 143	11.44% 100	33.64% 294	27.80% 243	10.76% 94	3.05
Sensible / Mature (5)	16.02% 140	10.30% 90	32.95% 288	30.09% 263	10.76% 94	3.10
Emotionally Responsive (6)	37.99% 332	30.78% 269	21.05% 184	8.35% 73	1.83% 16	2.05
Friendly (7)	9.38% 82	5.15% 45	19.68% 172	38.56% 337	27.35% 239	3.70

Curious (8)	46.57% 407	25.51% 223	19.34% 169	6.86% 60	1.83% 16	1.92
Compliant Flexible (9)	6.86% 60	13.62% 119	34.10% 298	35.70% 312	9.73% 85	3.28
Autonomous (10)	20.25% 177	24.60% 215	28.15% 246	20.02% 175	7.09% 62	2.69

Basic Statistics

	Minimum	Maximum	Median	Mean	Standard Deviation
Polite(1)	1.00	5.00	4.00	3.97	1.06
Sophisticated (2)	1.00	5.00	3.00	3.29	1.07
Moral (3)	1.00	5.00	2.00	2.07	1.19
Rational (4)	1.00	5.00	3.00	3.05	1.21
Sensible / Mature (5)	1.00	5.00	3.00	3.09	1.21
Emotionally Responsive (6)	1.00	5.00	2.00	2.05	1.04
Friendly (7)	1.00	5.00	4.00	3.69	1.19
Curious (8)	1.00	5.00	2.00	1.92	1.04
Compliant/ Flexible (9)	1.00	5.00	3.00	3.26	1.04
Autonomous (10)	1.00	5.00	3.00	2.69	1.20

Q10. How would you describe your virtual assistant in a couple of words / sentences?

Answered: 874 Skipped: 0

#	Responses	Date
1	Smart and really dumb at the same time	2018-02-07 17:43:05
2	Reliable. Modern. Looming. A machine of loving grace.	2018-02-07 17:08:19
3	A great help in many areas of my everyday life	2018-02-07 17:06:45
4	Incredibly useful	2018-02-07 16:57:04
5	Very useful for certain tasks	2018-02-07 16:51:44
6	not there yet but getting there, language seems to be the hurdle, also background noises. I have 10 echos in the house and 5 sonos and their atmospheres can clash sometime.	2018-02-07 16:50:33
7	The functionality is very similar to my cell phone's timer, web search and media apps. It feels like a convenient voice-activated input device, not a "personality" or "presence".	2018-02-07 16:48:34
8	A convenient device to have in your home	2018-02-07 16:33:19
9	useful	2018-02-07 15:55:28
10	a good hands free speaker. we are both learning the ropes	2018-02-07 15:48:55
11	It is a non-sentient machine with predetermined responses. It can't be autonomous, because it requires input from a human before it can do any action. It's very useful for everyday questions, updating me with current news, controlling certain aspects of my home, and playing music. It's still too early to use it for anything other than trivial requests.	2018-02-07 15:49:20
12	Extremely useful in daily life.	2018-02-07 15:35:51
13	its better than looking it up or doing it on my phone	2018-02-07 15:35:46
14	It's not perfect, but I appreciate how it's helped to automate the usage of other devices in my home (smart lights, Sonos speakers, TV), and it's always gaining features that make it even more useful.	2018-02-07 15:34:44
15	Utilitarian	2018-02-07 15:22:16
16	Very useful, but not as smart as it could be.	2018-02-07 15:13:07
17	Basic but developing	2018-02-07 15:02:09

18	Makes my life easier, for the most part. Not having to get out of bed to turn up the heat is life changing.	2018-02-07 15:03:18
19	Useful, makes my life a little easier, fun to use	2018-02-07 14:57:14
20	Useful	2018-02-07 14:54:32
21	2018-02-07 14:52:10
22	Convenient for some things, but hobbled by a lack of music streaming support in my country	2018-02-07 14:34:34
23	A machine, not a companion. At this point, the unintelligence of its behaviors means its just a really fancy pushbutton	2018-02-07 14:11:24
24	Handy tool for quick commands	2018-02-07 14:02:44
25	dumb weak useful	2018-02-07 13:59:24
26	Useful, luxury,	2018-02-07 13:59:28
27	Very handy for my whole family.	2018-02-07 13:55:21
28	Clunky. This is new technology and we're still not there yet. However, I am glad to be an early consumer.	2018-02-07 13:49:28
29	An added perk to life	2018-02-07 13:04:31
30	Compliant	2018-02-07 13:01:56
31	100% not human	2018-02-07 12:48:13
32	Very useful	2018-02-07 12:46:52
33	Useful	2018-02-07 12:41:31
34	Useful and gorgeous	2018-02-07 12:36:44
35	Great at Q&A, but needs more work for full communicative interaction. Still feels much more like an "assistant" than actual AI. It's meet my expectations for the price point.	2018-02-07 12:39:02
36	controls my home and music	2018-02-07 12:31:45
37	Sophisticated	2018-02-07 12:27:15
38	We love having it in the house and are finding new uses every day.	2018-02-07 12:26:32
39	My personal assistant	2018-02-07 12:17:24
40	Good	2018-02-07 12:20:52
41	Fucking awesome	2018-02-07 12:10:54
42	Handy to have	2018-02-07 12:09:23

43	Useful, however, full potential has not been realised yet I don't think.	2018-02-07 12:10:07
44	It's always there to help when I need it	2018-02-07 12:03:20
45	Useful toy	2018-02-07 12:02:25
46	Helpfull	2018-02-07 11:40:56
47	A good voice controlled music speaker, decent at controlling smart home devices, but getting a good answer to random questions is a hit and miss, still. (using German language)	2018-02-07 11:47:10
48	A good complement to our household.	2018-02-07 11:38:37
49	Getting there	2018-02-07 11:29:35
50	good, but could be better	2018-02-07 11:27:57
51	Great device.	2018-02-07 11:25:44
52	You don't know how much you need one until you have one.	2018-02-07 11:10:10
53	Idiot Savant	2018-02-07 10:56:53
54	Fun gadget	2018-02-07 10:56:41
55	Good but requires development of the AI to understand requests and communicate more naturally rather than respond to commands.	2018-02-07 10:55:58
56	useful	2018-02-07 10:53:55
57	Awesome	2018-02-07 10:42:28
58	New technology that works. Next step would have to be true AI.	2018-02-07 10:40:16
59	This question is far too open. Not sure what you mean by this.	2018-02-07 10:34:47
60	Non adaptive frustrating at times	2018-02-07 10:33:03
61	Great piece of work!	2018-02-07 10:24:42
62	Assistant	2018-02-07 10:24:59
63	Responsive, Effective and a good complement to my smartphone	2018-02-07 10:21:53

64	I think the fact that they answer in a proud way to insults, or inuendo questions almost makes them come across as superior, which i find quite tactful. I find it quite annoying that most companies default to female voices, especially given the technology is quite nascent. It's a bit demoralising to ask a women 4 times to play a song and receive the response - "sorry, I didn't quite get that". Also quite concerned with how will the generation that grows up around these devices will perceive women and interaction with people. Expecting stuff like "Mom! Fries!" ...it will be difficult to "unteach" this behaviors.	2018-02-07 10:24:33
65	Scripted	2018-02-07 10:12:39
66	It's not as smart as I expected but its fun	2018-02-07 10:05:11
67	Very useful indeed.	2018-02-07 10:00:55
68	not bad, needs improvement	2018-02-07 9:59:36
69	Convenient, novelty	2018-02-07 9:49:10
70	Convenient, helpful and a good way to keep track of timings handsfree such as snoozing alarms or checking on cooking.	2018-02-07 9:45:32
71	Best assistant ever!	2018-02-07 9:40:52
72	It does mostly what I ask	2018-02-07 9:41:19
73	Functional, but not as interactive as I would like	2018-02-07 9:40:33
74	Not very smart, good at a few routine tasks.	2018-02-07 9:17:07
75	Fun. Honestly the technology is further along than I expected. I wouldn't say I rely on it but its incredibly handy to have around.	2018-02-07 9:26:23
76	Really smart most of the time. Dumb as a rock on occasion.	2018-02-07 9:13:38
77	Does what I bought it for	2018-02-07 9:07:51
78	A machine	2018-02-07 9:01:25
79	A great bit of kit that helps with everyday life using smart technology. The device does not comprehend any rational or moral thoughts which i like	2018-02-07 9:04:49
80	Useful but unnecessary	2018-02-07 8:56:35
81	Great.	2018-02-07 8:46:16
82	A work in progress with the potential to be really great.	2018-02-07 8:22:03
83	Voice activated machine to deal with home automation	2018-02-07 8:24:21

84	Once you understand it's limitations, it is fairly useful.	2018-02-07 8:19:40
85	Fun, not the smartest	2018-02-07 8:20:11
86	Handy	2018-02-07 8:16:34
87	mostly useless	2018-02-07 8:13:25
88	A neat gimmick for people who dont know what to do with their money. Me included.	2018-02-07 8:12:43
89	N/A	2018-02-07 8:09:05
90	Sufficient and obedient. Streamlined and narrow	2018-02-07 8:11:02
91	Basic personality, mostly functional	2018-02-07 8:02:48
92	Handy	2018-02-07 7:58:42
93	Great at automating my home and performing common search queries	2018-02-07 7:59:56
94	I love them, but they are simply dumb! And its a heck of a deal to get it "wife" aprowed!	2018-02-07 7:54:03
95	Able to do basics conveniently and quickly, but nothing too complex or conversational (eg can't string multiple commands into one)	2018-02-07 7:44:39
96	Handy	2018-02-07 7:42:09
97	A nice toy with walled gardens	2018-02-07 7:40:01
98	A voice - activated weather and music box that doesn't know as much as I would like it to, yet.	2018-02-07 7:42:18
99	An integral part of our household, including my 4 year old	2018-02-07 7:38:30
100	Helpful every single day.	2018-02-07 7:38:17
101	it does what it was bought for	2018-02-07 7:25:04
102	My assistant	2018-02-07 7:24:51
103	useful, but limited	2018-02-07 7:23:15
104	Useful, but at the end of the day, it's just a speaker	2018-02-07 7:06:29
105	Terrific, could be better	2018-02-07 7:04:19
106	revolutionary	2018-02-07 7:01:43
107	Utility	2018-02-07 6:58:42
108	Whole house music biggest advantage	2018-02-07 7:00:06
109	a truly way to have smart homes	2018-02-07 6:59:53

110	Unfortunately limited in German, very utilitarian, little natural conversation	2018-02-07 6:56:50
111	Very useful for time management and asking questions.	2018-02-07 6:51:21
112	Helpful to what its intended for	2018-02-07 6:48:46
113	Programmed	2018-02-07 6:47:11
114	A great convenience, especially for smart home technologies.	2018-02-07 6:47:49
115	Home assistant	2018-02-07 6:46:39
116	Amazing	2018-02-07 6:46:05
117	Convenient	2018-02-07 6:40:19
118	Great for testing applications and exciting	2018-02-07 6:39:38
119	Google Home Mini is a great loud speaker, that can answer a million questions. I use it everyday	2018-02-07 6:33:06
120	Extremely useful but a little slow.	2018-02-07 6:27:09
121	virtual assistant	2018-02-07 6:19:27
122	Helpful. Entertaining. Learning.	2018-02-07 6:19:05
123	I love my smart speakers. Anything that makes my life easier. Reminders and timers are way easier to use than on a phone	2018-02-07 6:21:10
124	Smart assistant	2018-02-07 6:18:24
125	Growing	2018-02-07 6:10:29
126	We humans are getting more lazy because of these smart speakers, but I love the idea of the smart speaker.	2018-02-07 6:10:56
127	One of my greatest purchases	2018-02-07 6:09:57
128	A decent mix of convenience and feeling like I live in the future (Star Trek computer anyone?)	2018-02-07 6:06:20
129	Boop	2018-02-07 6:01:48
130	Friendly and great to have	2018-02-07 5:57:46
131	best clock/music speaker/weather forecast/light switch	2018-02-07 5:52:53
132	Helpful	2018-02-07 5:48:57
133	Cool concept that is frustratingly limited in certain degrees.	2018-02-07 5:46:35
134	Helpful	2018-02-07 5:45:29

135	Functional	2018-02-07 5:38:49
136	Hardware that provides an interface for Google's specified voice commands that works more reliably and with greater variation the user's natural spoken words than I generally expect it to.	2018-02-07 5:41:03
137	Indispensable when they work, impossibly frustrating when they don't.	2018-02-07 5:35:24
138	Handy devices that I keep finding more uses for. Can see them dominating the world, eventually.	2018-02-07 5:33:27
139	Useful	2018-02-07 5:31:00
140	Clever interface for connecting me to many useful functions	2018-02-07 5:28:16
141	Amazing	2018-02-07 5:25:58
142	Useful	2018-02-07 5:26:28
143	Capable of understanding most naturally spoken sentences	2018-02-07 5:27:32
144	The brain of a smart house. A constantly improving device that makes life easier, at the expense of some privacy	2018-02-07 5:24:21
145	Basic in today's day and age	2018-02-07 5:20:56
146	Utilitarian, just a tool, but it works well.	2018-02-07 5:16:38
147	It does what I pay for.	2018-02-07 5:16:56
148	Ok, so it's fine	2018-02-07 5:15:20
149	Does what I say as long as it's confined to what programmers want it to do. Works for me.	2018-02-07 5:14:34
150	Does what it was made to do	2018-02-07 5:13:47
151	Good	2018-02-07 5:13:15
152	Learning to do what it needs to do	2018-02-07 5:13:36
153	Convenient but not necessary	2018-02-07 5:09:17
154	It's a key part to my every day life	2018-02-07 5:10:13
155	It's a useful tool.	2018-02-07 5:11:14
156	It is useful for music and information. I also use it to broadcast messages throughout the home. I often use it as a timer.	2018-02-07 5:08:32
157	"This question requires an answer." No, it doesn't. :)	2018-02-07 5:10:46
158	useful	2018-02-07 5:08:39

159	A voice activated computer with an incredibly specific purpose	2018-02-07 5:06:38
160	Cool, neat	2018-02-07 5:04:54
161	It is the stepping stones to robot butlers.	2018-02-07 5:05:00
162	My google home works great for what I purchased it for: playing music.	2018-02-07 4:56:02
163	Great, but flawed. Makes me more frustrated than it should when things don't work as "designed"	2018-02-07 4:59:57
164	Great addition to my house, but probably not necessary. I really enjoy the ease of playing music through my devices, so this is a huge reason I use my smart assistants.	2018-02-07 4:57:06
165	"Sorry please say that again"	2018-02-07 4:51:05
166	Was more than I thought it would be. Use it for more talk as time goes on	2018-02-07 4:50:34
167	Exceeds expectations.	2018-02-07 4:52:44
168	Alexa is the queen of my domain. I pay her tribute by continually upgrading my home into a smart house.	2018-02-07 4:52:59
169	It's the Enterprise computer in a hockey puck.	2018-02-07 4:47:27
170	Helpful under most circumstances	2018-02-07 4:48:07
171	Very helpful	2018-02-07 4:45:18
172	Fun	2018-02-07 4:44:14
173	Amazing and smart	2018-02-07 4:44:04
174	It's a tool for controlling home automation, would be better if it had Google search support for answering questions	2018-02-07 4:42:20
175	Love it.	2018-02-07 4:40:11
176	Fun	2018-02-07 4:39:17
177	Great to see it develop. I love it	2018-02-07 4:38:23
178	Helpful	2018-02-07 4:38:16
179	Helpful and easy to use	2018-02-07 4:38:14
180	It's a fun gadget.	2018-02-07 4:39:01
181	It has been a welcomed addition to my lifestyle.	2018-02-07 4:39:29
182	As smart as a six grader and as likely to do what you want	2018-02-07 4:36:14
183	Shit with music selections but compliant.	2018-02-07 4:35:12

184	Idiotic. Annoying as google and amazon do not play nice with each other.	2018-02-07 4:34:28
185	A basic home helper/butler. Reminders, turning on/off lights, telling you the day's news.	2018-02-07 4:34:08
186	Fun, handy, interesting, futuristic	2018-02-07 4:32:22
187	utilitarian	2018-02-07 4:28:29
188	Ever ready to attempt my request.	2018-02-07 4:29:21
189	Helpful	2018-02-07 4:28:05
190	Great helper around the house	2018-02-07 4:26:58
191	Limited by lack of vision	2018-02-07 4:27:24
192	It turns my lights on and off without me going to the light switch, which is genuinely worth the set up and cost to me. I don't really care what else it does -- if it only did that, I'd be equally happy.	2018-02-07 4:27:15
193	Very helpful and easy to use.	2018-02-07 4:28:18
194	Audible remote control	2018-02-07 4:25:07
195	Useful. Helpful. Helps me look at my phone less.	2018-02-07 4:23:17
196	My Google Home Mini quickly accesses information already available on the Internet. It does so quickly and easily and does a good job understanding my commands. I would even say that it is fun to use; I enjoy seeing what it can do and seeing how new commands and features have been added over time.	2018-02-07 4:23:57
197	Useful	2018-02-07 4:20:58
198	Useful	2018-02-07 4:19:07
199	Our echo is very convenient to use in our apartment (especially in the kitchen).	2018-02-07 4:20:37
200	Great piece of tech	2018-02-07 4:18:06
201	ties my house together	2018-02-07 4:15:08
202	When it works, it's great, but when it fails, it's useless and basically as good as a paperweight	2018-02-07 4:14:50
203	Works OK	2018-02-07 4:14:19
204	It's a electronic device that helps me achieve some tasks and helps me organize my day	2018-02-07 4:16:38
205	Integral part of my life now	2018-02-07 4:12:26
206	It's a great addition for fun and turning on the lights	2018-02-07 4:12:34

207	It's extremely convenient	2018-02-07 4:11:01
208	Great at simple commands but struggles when things get too specific/wordy.	2018-02-07 4:11:19
209	Usually does what it's designed to do.	2018-02-07 4:09:53
210	My Google homes have become part of life in my house. It has encouraged me to pursue other home automation products because of the extensive integrations offered through Google Assistant.	2018-02-07 4:10:52
211	Easy to use, informative	2018-02-07 4:05:29
212	A search engine with a voice	2018-02-07 4:04:38
213	Useful	2018-02-07 4:03:10
214	Does what I want, although I wish it were capable of more	2018-02-07 4:12:20
215	Great but without smart home devices, it doesn't do that much	2018-02-07 4:05:13
216	Fun	2018-02-07 4:00:56
217	Neat	2018-02-07 3:58:41
218	Convenient enough to outweigh any privacy concerns	2018-02-07 3:57:16
219	A very helpful shadow of an AI.	2018-02-07 3:56:52
220	Very helpful when needed	2018-02-07 3:55:52
221	Helpful	2018-02-07 3:53:20
222	Entertaining, informative	2018-02-07 3:52:44
223	Robotic	2018-02-07 3:53:05
224	Good for day to days tasks but has more potential	2018-02-07 3:52:01
225	Cool, new age	2018-02-07 3:51:47
226	It gets better with every update.	2018-02-07 3:51:38
227	Helpful but limited	2018-02-07 3:50:14
228	Sometimes annoying.	2018-02-07 3:45:40
229	Good! But I want to string commands.	2018-02-07 3:47:20
230	Dumb	2018-02-07 3:44:23
231	The future! These will be used all over homes and cars within the decade.	2018-02-07 3:42:22
232	A useful gadget.	2018-02-07 3:52:41

233	Learning	2018-02-07 3:37:35
234	Helpful for managing my large family. Great at managing other smart devices.	2018-02-07 3:38:44
235	My dot has been pretty useful so far. I plan to get my family some so we can communicate easily.	2018-02-07 3:36:56
236	Easy to use	2018-02-07 3:34:22
237	Worthwhile	2018-02-07 3:32:13
238	Highly useful and occasionally infuriating.	2018-02-07 3:31:13
239	A wonderful interface to access technology with voice.	2018-02-07 3:32:28
240	Handy for smart home control; long way to go in terms of A.I.	2018-02-07 3:32:24
241	na na na na	2018-02-07 3:31:59
242	Useful	2018-02-07 3:31:17
243	Useful tool	2018-02-07 3:30:22
244	Annoying, used to hear well, is now going deaf. Planned obscelence? Confuses it's self with words, wether or not speaking clearly with or without an accent and slang	2018-02-07 3:31:26
245	Impressive	2018-02-07 3:29:10
246	Personal secretary	2018-02-07 3:28:49
247	It is a tool	2018-02-07 3:23:02
248	Great for playing music, kids love it	2018-02-07 3:21:18
249	Helpful, pleasant, happy.	2018-02-07 3:28:47
250	Novelty	2018-02-07 3:21:15
251	Useful, necessary, fun	2018-02-07 3:19:23
252	Very useful, but lacking continuity in volume and command controls.	2018-02-07 3:20:35
253	It does what my phone's assistant does, but I like that I can use it to control Netflix and other things on my TV.	2018-02-07 3:18:36
254	A box that can search the internet	2018-02-07 3:17:48
255	Great Home assistant to automate things around the house, great for listening to music inside and outside with voice commands.	2018-02-07 3:15:14
256	Useful	2018-02-07 3:14:38
257	It's great.	2018-02-07 3:38:48

258	Could be better with updates the community wants	2018-02-07 3:15:54
259	Ehehe	2018-02-07 3:14:49
260	Useful	2018-02-07 3:12:38
261	helpful for simple tasks	2018-02-07 3:12:09
262	Ernest but limited	2018-02-07 3:14:08
263	Helpful	2018-02-07 3:09:02
264	Basic tasks, but more complex tasks or string a few together and it doesn't work	2018-02-07 3:07:56
265	Still not smart enough	2018-02-07 3:08:22
266	A wonderful purchase, a great addition to my household. I have bought two more since my initial Google Home purchase. The Google Home is better than Amazon Echo, I have/had a few of those and they do not seem as robust or knowledgeable.	2018-02-07 3:07:44
267	Useful, voice input for things not autonomous.	2018-02-07 3:05:19
268	Speaker device that responds to voice commands	2018-02-07 3:04:18
269	It's just a tool: like another appliance. Did I need it? No, but it's fun to have.	2018-02-07 3:05:08
270	Google Assistant keeps growing, and I am excited for the future.	2018-02-07 3:02:16
271	best radio ever	2018-02-07 3:02:18
272	Quickly integrated into my daily life	2018-02-07 3:03:31
273	Smart, but distinctly robotic. Long way to go, before it could understand and predict things to do	2018-02-07 3:01:33
274	Great addition to a home or office space for fast and easy information.	2018-02-07 3:01:25
275	Useful, fun, but probably not essential.	2018-02-07 2:58:18
276	Nice addition to an increasingly smart home.	2018-02-07 2:59:21
277	Convinient	2018-02-07 2:51:30
278	Useful voice remote	2018-02-07 2:50:43
279	It's okay	2018-02-07 2:48:49
280	She is a bloody mad men	2018-02-07 2:49:27
281	Very convenient and helpful	2018-02-07 2:49:37
282	Just about there.	2018-02-07 2:48:29

283	Helpful.	2018-02-07 2:47:32
284	Nice gimmick but long way to go	2018-02-07 2:45:47
285	Perfection from Google!	2018-02-07 2:44:03
286	I mostly use it to play music and control all my lights. There really isn't much I would change.	2018-02-07 2:43:24
287	Handy, but not vital to my daily routine.	2018-02-07 2:43:02
288	Neat with lots of kinks to work out.	2018-02-07 2:39:49
289	Helpful and a novelty.	2018-02-07 2:42:59
290	Super easy to set up and use.	2018-02-07 2:37:49
291	A step towards assistant but not there yet	2018-02-07 2:37:19
292	A tool to make life easier.	2018-02-07 2:36:59
293	It's good for controlling my lights	2018-02-07 2:37:30
294	It serves a purpose but needs improvement.	2018-02-07 2:36:34
295	Useful	2018-02-07 2:34:08
296	Friendly, happy.	2018-02-07 2:33:35
297	Very much a "magic word" recognizing machine. Google's AI is pretty good at figuring out what I want, but it does require some figuring out. Had enough capabilities that it is pretty useful once I've figured out the "magic words" so to speak.	2018-02-07 2:35:30
298	Simple, intuitive, fun	2018-02-07 2:34:26
299	Useful	2018-02-07 2:30:40
300	Smart, Helpful	2018-02-07 2:30:16
301	Always learning	2018-02-07 2:27:24
302	Functional	2018-02-07 2:26:44
303	made by google!	2018-02-07 2:23:54
304	Much better than Siri but still needs a lot of work.	2018-02-07 2:25:33
305	Just a machine	2018-02-07 2:22:16
306	Useful and efficient	2018-02-07 2:21:36
307	A tool.	2018-02-07 2:17:32
308	Works well, google knows how users use.	2018-02-07 2:16:42

309	Convenient	2018-02-07 2:12:23
310	Useful, practical, time saving.	2018-02-07 2:12:17
311	Always willing to help	2018-02-07 2:09:50
312	Good at following instructions if they're spoken exactly as expected. Rigid and inflexible in most cases.	2018-02-07 2:08:37
313	Needs the ability to read web pages or search results	2018-02-07 2:04:40
314	A very fun and pretty useful tool	2018-02-07 2:02:19
315	handy and convenient	2018-02-07 2:05:37
316	A value to the home that is near perfection, but has a few bugs to be worked out.	2018-02-07 1:59:33
317	Useful, especially as a hub for a connected home.	2018-02-07 1:57:12
318	Handy	2018-02-07 1:56:32
319	Useful, but need improvement	2018-02-07 1:55:09
320	They're not people, you know. They might be in the future, I suppose.	2018-02-07 1:55:18
321	While a helpful device, it's mostly an alarm clock and light switch. The echo isn't great at answering questions unless they're very specific or general.	2018-02-07 1:55:05
322	Helpful	2018-02-07 1:53:08
323	Always ready, always helpful to help with commands.	2018-02-07 1:54:36
324	Decent	2018-02-07 1:51:26
325	Wonderful, when it works	2018-02-07 1:56:55
326	Futuristic	2018-02-07 1:49:54
327	Its ok	2018-02-07 1:49:26
328	Nice gadget I have 6 of them, but still a work in progress but moving forward at a good pace.	2018-02-07 1:53:01
329	helpful servant	2018-02-07 1:47:36
330	It's a \$30 timer.	2018-02-07 1:48:33
331	A device that works well at what it was meant to do.	2018-02-07 1:44:44
332	Gets the job done	2018-02-07 1:42:33
333	I'm surprised every day both by the things my Google home can understand and the things it can't understand	2018-02-07 1:41:36

334	There is both Google Home, and Alexa, and I much prefer Google Home's understanding, realistic nature, and less robotic voice (at times)	2018-02-07 1:43:16
335	cool	2018-02-07 1:40:39
336	Helpful and playful when needs be.	2018-02-07 1:39:35
337	Functional idiot	2018-02-07 1:39:07
338	It's a handy tool to have, and really convenient.	2018-02-07 1:40:32
339	No typing	2018-02-07 1:35:15
340	It is a smart tool. One that I use to make ordinary tasks easier.	2018-02-07 1:34:48
341	Acceptable	2018-02-07 1:34:39
342	Useful, but still limited	2018-02-07 1:33:02
343	I am still pretty amazed at the level of comprehension to voice commands. Even when soft spoken, my wishes are interpreted correctly. It is quickly becoming indispensable.	2018-02-07 1:35:47
344	Effective	2018-02-07 1:32:47
345	It works well, but much better in the USA	2018-02-07 1:36:24
346	The start of something big, constantly being updated with new features and getting better at its job almost daily. It is the reason I am getting more smart home devices.	2018-02-07 1:34:10
347	fun handy	2018-02-07 1:28:29
348	Convenient, easy to use	2018-02-07 1:28:47
349	Allows increased convenience in several ways: music, lights, general knowledge questions, lists, etc.	2018-02-07 1:24:24
350	Cool	2018-02-07 1:16:54
351	Good at what it's suppose to do with some bonuses.	2018-02-07 1:15:27
352	Basic appliance at this point. Just a verbal replacement for my phone google assistant functions, that sometimes works.	2018-02-07 1:14:46
353	Convenient, like how it works in the background	2018-02-07 1:12:27
354	smart	2018-02-07 1:09:01
355	A handsfree tool with specific, limited capabilities that can be repeated with a high degree of accuracy. Occasionally, it's surprisingly useful.	2018-02-07 1:10:15
356	Great	2018-02-07 1:08:55

357	My assistant has trained me how to work with it. I have a head last success training it to work with me	2018-02-07 1:08:19
358	You need to have an NA for question 9... My speakers do not have feelings	2018-02-07 1:06:00
359	Useful and informative. Still working on how to communicate with it.	2018-02-07 1:04:43
360	Still improving	2018-02-07 1:01:23
361	Does some neat things, still needs work.	2018-02-07 1:01:32
362	Helpful, always evolving, fun, intuitive	2018-02-07 0:58:45
363	Useful and convenient	2018-02-07 0:56:50
364	More handy every day	2018-02-07 0:59:42
365	A joy to interact with	2018-02-07 0:54:08
366	Helpful	2018-02-07 1:02:47
367	Google stopped developing for it so it feels like it's getting worse.	2018-02-07 0:50:15
368	Useful	2018-02-07 0:50:07
369	Practical	2018-02-07 0:45:34
370	I love it!	2018-02-07 0:46:47
371	Useful but limited	2018-02-07 0:44:13
372	Easy to use and helpful	2018-02-07 0:41:50
373	It gets the job done. I expected less and got more.	2018-02-07 0:50:00
374	It's a piece of electronic equipment. It does what I ask it to do.	2018-02-07 0:41:27
375	Useful	2018-02-07 0:41:02
376	Can't live without it. Use it for home automation and it has made me lazy...	2018-02-07 0:40:34
377	I like it	2018-02-07 0:40:10
378	Is a companion more than a friend.	2018-02-07 0:37:17
379	Smart. Funny. Useful.	2018-02-07 0:34:28
380	2018-02-07 0:31:47
381	convenient	2018-02-07 0:32:21
382	pretty sweet	2018-02-07 0:27:14
383	Handy	2018-02-07 0:27:09

384	Good value for money.	2018-02-07 0:26:46
385	It's great.	2018-02-07 0:25:09
386	In early stages	2018-02-07 0:27:04
387	Amazin	2018-02-07 0:24:29
388	Useful	2018-02-07 0:21:02
389	Useful for basic tasks	2018-02-07 0:18:42
390	Helpful tool, but doesn't meet all needs perfectly.	2018-02-07 0:18:12
391	it does what i need it to do	2018-02-07 0:16:50
392	Helpful, convenient, simple to use	2018-02-07 0:15:56
393	Helpful and fun	2018-02-07 0:17:25
394	I don't use it much but I love it	2018-02-07 0:16:04
395	A wonderful first step. I don't think it'll be possible to retrofit the upcoming more empathic-realialistic interactions into the older models, so I will probably replace her, then.	2018-02-07 0:19:20
396	Revolutionary as the technology evolves, particularly with more prevalent presense and the availability to connect to varying screens	2018-02-07 0:21:12
397	Useful for the purpose of controlling a smart house with your voice.	2018-02-07 0:13:35
398	It's a wonderful unpaid assistant, it accompanies me as I work and aids in solving questions and scheduling.	2018-02-07 0:10:41
399	An expensive light switch.	2018-02-07 0:11:12
400	Usable.	2018-02-07 0:15:16
401	pretty cool	2018-02-07 0:09:35
402	Useful, but in need of more functionality and better understanding of requests	2018-02-07 0:07:41
403	The start of a new hobby.	2018-02-07 0:06:46
404	Useful tool, helpful device	2018-02-07 0:08:50
405	It is useful. It is not an artificial intelligence quite yet.	2018-02-07 0:04:33
406	A really great home help for lights, Chromecast. Looking forward to new features in UK soon. Definitely going to buy a mini so I don't have to raise my voice in another room!	2018-02-07 0:05:08
407	My personal secretary	2018-02-07 0:08:52

408	Convenient home automation plus music	2018-02-07 0:05:06
409	A new way to interact with iot or stuff	2018-02-07 0:01:39
410	Objective and politically correct	2018-02-06 23:59:58
411	Great tool and entertainment device	2018-02-07 0:02:11
412	Virtual Intelligence not artificial intelligence	2018-02-06 23:56:27
413	Improving	2018-02-06 23:55:31
414	It makes simple tasks much more convenient to perform.	2018-02-06 23:47:30
415	Amazing as a first step	2018-02-06 23:42:34
416	Love it. Love video intercom. Music and radio are important to me	2018-02-06 23:45:39
417	Just waiting for AI to catch up to wants and needs. For what it is it's fine.	2018-02-06 23:43:07
418	need to know how to handle -sometimes buggy (non reliable answers)	2018-02-06 23:41:39
419	Does what it knows well, but could be expanded upon	2018-02-06 23:41:44
420	She's nice. Very cordial	2018-02-06 23:41:05
421	N/A	2018-02-06 23:38:16
422	Helpful	2018-02-06 23:38:17
423	Very helpful	2018-02-06 23:37:55
424	Needs more apps	2018-02-06 23:34:49
425	Convenient	2018-02-06 23:33:34
426	.aaaaaaa	2018-02-06 23:33:06
427	Essential but flawed, expect platform to improve over next few years	2018-02-06 23:34:26
428	Alexa is very useful for playing music and controlling smart devices	2018-02-06 23:32:53
429	Useful, entertaining, informative	2018-02-06 23:30:59
430	Verbose	2018-02-06 23:30:00
431	Helpful more often than not	2018-02-06 23:26:58
432	Promising	2018-02-06 23:26:26
433	I love it	2018-02-06 23:26:24
434	Robot lady who lives in my house.	2018-02-06 23:21:32

435	It replaces several devices with just the one	2018-02-06 23:21:15
436	useful	2018-02-06 23:16:18
437	Meets my needs and makes daily living better	2018-02-06 23:16:33
438	It's a good start towards something better.	2018-02-06 23:16:04
439	I've gotten to admit it's getting better, it's getting better all the time	2018-02-07 1:16:15
440	none	2018-02-06 23:14:06
441	Not quite there yet	2018-02-06 23:15:50
442	Love the gadget	2018-02-06 23:14:13
443	automated home butler	2018-02-06 23:12:33
444	extremely smart but not very useful in general for me	2018-02-06 23:13:00
445	At times frustrating but generally functional. New features slow to arrive in UK.	2018-02-06 23:13:03
446	Informational Music player	2018-02-06 23:11:34
447	Does basic tasks well, but has lots of room for potential.	2018-02-06 23:04:21
448	Convenient, fun to play with	2018-02-06 23:03:45
449	useful	2018-02-06 23:04:16
450	Thumbs up emoji	2018-02-06 23:02:26
451	very useful	2018-02-06 23:02:49
452	Useful	2018-02-06 23:00:10
453	It's a nice thing to play some music ,set some alarms etc. Without using hands	2018-02-06 23:04:20
454	Natural	2018-02-06 23:00:07
455	It's helpful. I'm disabled and it turns the lights on for me.	2018-02-06 22:58:24
456	Fun	2018-02-06 22:56:52
457	It's a great start to having an actual smart home	2018-02-06 22:57:06
458	Glimpse of the future.	2018-02-06 22:54:14
459	A device that doesn't put questions with compulsory text answers in a survey.	2018-02-06 22:52:54
460	Accessible, convenient	2018-02-06 22:53:04

461	We are at the beginning of this technology development. The assistant can only be described as the best we've seen but so very primitive at the same time.	2018-02-06 22:52:52
462	This is the future	2018-02-06 22:51:01
463	Useful	2018-02-06 22:52:38
464	Useful	2018-02-06 22:48:23
465	Getting there!	2018-02-06 22:46:32
466	Pretty cool but don't use it as much as I want to as I haven't purchased any smart home devices yet.	2018-02-06 22:47:26
467	nifty, but need a few years until they're really there	2018-02-06 23:56:59
468	Fun	2018-02-06 22:41:22
469	Useful	2018-02-06 22:40:38
470	Intelligent but verbally dumb	2018-02-06 22:41:14
471	not applicable	2018-02-06 22:40:04
472	Getting there	2018-02-06 22:37:55
473	..	2018-02-06 22:38:07
474	Useful, but not completely reliable	2018-02-06 22:37:35
475	Useful tool	2018-02-06 22:36:32
476	Kids love it	2018-02-06 22:36:28
477	a useful tool	2018-02-06 22:36:27
478	Helpful and useful	2018-02-06 22:37:23
479	Robotic best friend	2018-02-06 22:34:18
480	Polite, Compliant	2018-02-06 22:33:32
481	It is a great automation tool for my home. It also provides many QOL services such as netflix casting and being able to play music/podcasts.	2018-02-06 22:30:43
482	Not essential but a huge time saver for the everyday.	2018-02-06 22:30:30
483	Fun	2018-02-06 22:30:46
484	I don't want to type	2018-02-06 22:28:26
485	Helpful	2018-02-06 22:25:22
486	useful	2018-02-06 22:22:01

487	N/A	2018-02-06 22:21:35
488	A beloved member of the household	2018-02-06 22:21:43
489	A speaker I didn't need before I bought it and I wouldn't replace it if someone stole it or it broke.	2018-02-06 22:19:42
490	Can be improved	2018-02-06 22:19:56
491	A voice interface for the internet.	2018-02-06 22:37:15
492	There to help	2018-02-06 22:16:57
493	A useful addition to the home. Capable of more	2018-02-06 22:15:40
494	Smart, responsive, ready	2018-02-06 22:11:19
495	First gen of its kind	2018-02-06 22:09:50
496	It makes certain tasks more convenient, but can't do everything.	2018-02-06 22:08:26
497	Mainly just a music speaker for me. Lots of things it can't do yet.	2018-02-06 22:09:42
498	Stupid technology to make my life more efficient	2018-02-06 22:06:02
499	Useful tool	2018-02-06 22:05:44
500	Gives Intelligent feedback	2018-02-06 21:55:27
501	Functional but not friendly like a human	2018-02-06 21:55:25
502	Useful	2018-02-06 21:54:22
503	Good	2018-02-06 21:52:21
504	Useful for known commands, but hopeless for understanding context or natural language.	2018-02-06 21:52:40
505	Very useful technology with a lot of promise.	2018-02-06 21:54:07
506	Good. Could be better	2018-02-06 21:51:03
507	Helpful. Useful	2018-02-06 21:50:38
508	love it	2018-02-06 21:43:39
509	Limited until the technology progresses	2018-02-06 21:42:51
510	Really cool	2018-02-06 21:48:52
511	Made me even lazier	2018-02-06 21:38:04
512	Sometimes a little too verbose. Frustrating when it doesn't understand because I've said something in a slightly different way than it would prefer.	2018-02-06 21:38:15

513	More or less a voice controlled keyboard to Google.	2018-02-06 21:36:50
514	More convenient then picking up my phone for some things.	2018-02-06 21:36:36
515	Honestly, its cool, but with hindsight i wouldnt get one again now. I dont think the technology is quite where it needs to be at the moment. I am english but my partner is scandinavian. Its handling of accents is pretty awful, and when it comes down to it, its essentially a voice controlled music unit until they get it to interact with Microsoft for me.	2018-02-06 21:32:43
516	Limited but functional	2018-02-06 21:32:25
517	Easy to use, hard to master	2018-02-06 21:30:49
518	Life changing... slowly	2018-02-06 21:28:50
519	Useful but not quite a necessity yet	2018-02-06 21:26:03
520	Voice recognition coupled with making api calls	2018-02-06 21:27:33
521	Useful	2018-02-06 21:25:15
522	Extremely useful	2018-02-06 21:22:28
523	Helpful	2018-02-06 21:25:22
524	Useful alternative to phone based assistant	2018-02-06 21:22:46
525	It's a tool so that I don't have to pull out my cell phone to do home automation stuff. Assigning any sort of personality or humanity to it is foolish.	2018-02-06 21:23:48
526	Useful	2018-02-06 21:18:05
527	Integral part of our smart home	2018-02-06 21:14:28
528	useful	2018-02-06 21:12:46
529	Potentially cool.	2018-02-06 21:12:15
530	Very useful and surprisingly humanlike	2018-02-06 21:11:26
531	Convenient	2018-02-06 21:05:00
532	Very useful, but still feels like putting in button/mouse commands when using it, rather than a natural flow of commands in talking	2018-02-06 21:03:19
533	I wish higher end speakers were made for gh	2018-02-06 20:59:37
534	A computer algorithm designed to answer questions and needs. I do not see it as deserving of empathy the way I would an animal.	2018-02-06 20:57:32
535	Acceptable	2018-02-06 20:56:11

536	useful	2018-02-06 20:52:35
537	pretty good at turning things on and off	2018-02-06 20:53:04
538	Functional.	2018-02-06 20:50:24
539	Great and rage inducing but Great	2018-02-06 20:51:47
540	Smart speaker PLUS	2018-02-06 20:48:20
541	Basic but effective tool	2018-02-06 20:47:03
542	Useful tool for a smart home	2018-02-06 20:48:28
543	an interactive tool	2018-02-06 20:46:26
544	A very expensive paper weight	2018-02-06 20:45:29
545	nonessential convenience	2018-02-06 20:41:44
546	Has the appearance of being smart whilst not fully understanding the nuances of human language	2018-02-06 20:41:41
547	Inanimate	2018-02-06 20:45:21
548	Quaint	2018-02-06 20:37:03
549	Useful. And better Natural language processing than Alexa. Still needs some improvement.	2018-02-06 20:33:35
550	it works ok	2018-02-06 20:37:09
551	A glorified speaker	2018-02-06 20:31:25
552	Efficient, helpful, mainstream	2018-02-06 20:27:22
553	Easy to use, understands a bit better without the finger poke command prompt (feels like I'm the sergeant major and it's the new recruit sometimes)	2018-02-06 20:28:39
554	I enjoy having it!	2018-02-06 20:24:31
555	A tool, just like any other tool - Useful in the right context. Good if you know what to expect	2018-02-06 20:23:11
556	Useful. Informative	2018-02-06 20:26:00
557	The Google home is great at controlling smart devices but sometimes doesn't hear	2018-02-06 20:22:07
558	Having potential to be an actual "personal assistant" but needs a more open platform and more time put into natural language processing.	2018-02-06 20:24:29
559	Very handy	2018-02-06 20:21:58
560	It's the best Bluetooth speaker I've ever had.	2018-02-06 20:21:07
561	It is great	2018-02-06 20:19:15

562	A helpful, but unnecessary, tool to simplify home commands	2018-02-06 20:19:53
563	Politely intelligent	2018-02-06 20:19:56
564	Amazing	2018-02-06 20:18:34
565	Disappointinh. Not AI. Handy music player. The start of the future and the build up of exciting things to come.	2018-02-06 20:21:26
566	It is my whole home music system.	2018-02-06 20:18:17
567	Internet connected speaker that sits on a shelf.	2018-02-06 20:16:07
568	Visiting places without one feels like traveling back in time. This is what I imagined as the future when I was a child.	2018-02-06 20:17:02
569	Useful	2018-02-06 21:28:59
570	Lacking the AI I would have expected.	2018-02-06 20:14:37
571	Assistive	2018-02-06 20:14:23
572	A tool	2018-02-06 20:13:43
573	nice addition to home automation	2018-02-06 20:16:06
574	Frustrating, informative, pleasant	2018-02-06 20:20:26
575	Amazing	2018-02-06 20:11:28
576	Useful, somewhat life-like. Getting closer to real potential	2018-02-06 20:11:20
577	It's great in the kitchen and as an alarm clock, has great sound quality and hears my voice clearly	2018-02-06 20:11:15
578	Convenient and user friendly	2018-02-06 20:09:27
579	Useful AF	2018-02-06 20:09:30
580	Makes my life slightly more convenient, but has a very long way to go. GH is very bad at hearing me over music and picks up my voice better when I speak in lower tones, which is HIGHLY annoying.	2018-02-06 20:10:22
581	Lots of potential	2018-02-06 20:12:57
582	My little helper	2018-02-06 20:08:34
583	useful	2018-02-06 20:07:12
584	Useful for small things like playing music and setting timers but clunky for watching/controlling video	2018-02-06 20:20:59
585	It does what it is designed for	2018-02-06 20:06:45
586	Reliable and convenient	2018-02-06 20:06:31

587	A "Smart Speaker" and a "Virtual Assistant" are two VERY different things. Smart speakers don't have to have a Virtual Assistant. And how do you define "Smart Speaker?" I really hope this is just a preliminary survey and you will re-work some of these questions. They are all super subjective (i.e. "How APPLICABLE are the following...")	2018-02-06 20:10:55
588	A valuable resource	2018-02-06 20:01:59
589	It doesn't listen sometimes.....	2018-02-06 19:59:49
590	convenient but limited	2018-02-06 19:58:35
591	Still often frustrating, but very useful	2018-02-06 20:00:00
592	A tool	2018-02-06 19:55:14
593	An useful addition, but not the whole story	2018-02-06 19:55:43
594	not very human-like	2018-02-06 19:51:36
595	Helpful. Very Helpful/	2018-02-06 19:49:33
596	It's a nice gadget but it needs to be more flexible	2018-02-06 19:51:53
597	Bit of a gimmick	2018-02-06 19:49:56
598	my personal assistant that sometimes misunderstands me	2018-02-06 19:47:40
599	A useful tool that's quicker than a phone.	2018-02-06 19:47:19
600	Very good in some tasks, very poor in others	2018-02-06 19:46:18
601	discrete looking, plays music well, replaces the stereo	2018-02-06 19:43:35
602	It's smart for something but not in other aspects. Can't do everything I want and have to be very precise with my commands.	2018-02-06 19:42:21
603	nongraphical UI.	2018-02-06 19:40:45
604	I'm hooked. Doubt I will live with out one now.	2018-02-06 19:43:41
605	Highly useful and convenient for quick info	2018-02-06 19:38:51
606	Functional but unable to show politeness (ie; "ok Google, turn off the lights" "ok, lights turning off" "thanks Google" -no response. This is disappointing, because I have 4 kids who could be learning additional times to express politeness.	2018-02-06 19:38:07
607	Very Useful	2018-02-06 19:34:24
608	Handy, fun, easy to use	2018-02-06 19:34:43
609	Useful	2018-02-06 19:33:17

610	Useful adaptive	2018-02-06 19:34:50
611	A helpful toy.	2018-02-06 19:32:30
612	It's cool technology that is better than I hoped, but still not perfect.	2018-02-06 19:31:00
613	Helpful but limited	2018-02-06 19:28:37
614	What it is supposed to be	2018-02-06 19:29:16
615	Convenient and accessible	2018-02-06 19:29:31
616	A time and effort saving convenience	2018-02-06 19:30:08
617	Decent home automation hub	2018-02-06 19:26:19
618	Unmature prototype	2018-02-06 19:26:52
619	Helpful	2018-02-06 19:22:07
620	Helpful.	2018-02-06 19:21:46
621	It is a convenient addition to my life and I have been using it more the longer I've owned it as its skills improve.	2018-02-06 19:22:05
622	Not as clever as I hoped. Less clever than just typing into Google.	2018-02-06 19:21:57
623	The Echo is a task-master only doing what it has the skills for, the Home is a partner willing to try to find a solution to so many more things you request from it	2018-02-06 19:18:31
624	It's a system that's good at pretending to be human until it's not	2018-02-06 19:18:26
625	Awesome!	2018-02-06 19:54:15
626	Useful, novel, fun	2018-02-06 19:18:06
627	The Google Home acts as judge and jury in all bets between my wife and I.	2018-02-06 19:19:46
628	Useful but lacking in quite a few areas.	2018-02-06 19:13:29
629	Convenient, easy to use	2018-02-06 19:14:38
630	It's awesome.	2018-02-06 19:11:30
631	Best thing I've bought	2018-02-06 19:10:49
632	Very convenient addition to my home	2018-02-06 19:17:07
633	Very neat	2018-02-06 19:08:28
634	A fantastic assistant to my every day life	2018-02-06 19:09:06
635	Beta	2018-02-06 19:10:55

636	okay for \$30	2018-02-06 19:08:08
637	It performs functionally.	2018-02-06 19:09:03
638	Useful	2018-02-06 19:05:17
639	useful, but sometimes unexpectedly limited.	2018-02-06 19:05:22
640	Its like a pet.	2018-02-06 19:05:23
641	A useful tool.	2018-02-06 19:40:59
642	Very useful	2018-02-06 19:05:23
643	Does what I ask it	2018-02-06 19:02:55
644	Surprisingly useful, unintentionally frustrating, overall novel but not yet essential	2018-02-06 19:01:49
645	great addition to a smart home.	2018-02-06 19:01:09
646	Humorous	2018-02-06 18:59:11
647	A useful addition	2018-02-06 18:56:58
648	Convenient	2018-02-06 18:57:07
649	convenient luxury	2018-02-06 18:54:52
650	Great	2018-02-06 19:29:57
651	A handy tool	2018-02-06 18:51:41
652	Easy to use machine	2018-02-06 18:49:48
653	Almost there. It would be good to actually have a conversation with it, getting it to understand what I want without certain commands and having them be exact... le turn on bathroom light. But I still enjoy it and how easy some things are to do with it.	2018-02-06 18:52:57
654	Getting better with each update	2018-02-06 18:50:29
655	Convenient, useful, and inexpensive.	2018-02-06 18:50:36
656	It does several things well, but I wish it did more.	2018-02-06 18:48:39
657	NANANANANA	2018-02-06 18:48:02
658	Alexa is a GREAT tool, and I cannot wait for the day VI's are even easier to have a conversation with.	2018-02-06 18:48:20
659	Useful when working as expected	2018-02-06 18:46:34
660	Useful	2018-02-06 18:46:01
661	useful and convenient	2018-02-06 18:44:29

662	Helpful and frustrating	2018-02-06 18:45:20
663	She tries her best to do what I ask	2018-02-06 18:42:10
664	It does what I need it to do most of the time.	2018-02-06 18:42:58
665	Smart but limited	2018-02-06 18:41:20
666	Great device fo whole family	2018-02-06 18:38:40
667	Great gimmick, needs more time to take over	2018-02-06 18:40:49
668	Great fun, useful too.	2018-02-06 18:37:40
669	Useful for its limited purpose	2018-02-06 18:36:56
670	Useful	2018-02-06 18:34:45
671	The first iteration of what I'm sure will look like very basic in a few years time.	2018-02-06 18:34:43
672	A great utility to have	2018-02-06 18:35:11
673	It's a useful tool.	2018-02-06 18:38:01
674	She's a very nice light switch with music and weather to boot	2018-02-06 18:50:46
675	Very helpful and streamlined. Easy to learn and understand. Does not understand you if you don't enunciate	2018-02-06 18:34:13
676	I love it!	2018-02-06 18:35:05
677	Useful	2018-02-06 18:36:56
678	Great tool that will be in more and more homes every time you look	2018-02-06 18:31:24
679	Often smart, sometimes dumb.	2018-02-06 18:30:02
680	The Google Assistant voice/speech synthesis is pleasant to interact with because it sounds so natural. It's just fun having Google in a speaker.	2018-02-06 18:36:35
681	Good. Works reasonably well most of the time. Sometime annoying when it does not work	2018-02-06 18:30:03
682	Nice and polite but not very smart. Doesn't yet feel like speaking to a "artificial intelligence"as a lot of things arent understood	2018-02-06 18:33:52
683	The future	2018-02-06 18:28:00
684	unpredictable	2018-02-06 18:27:15
685	Good! Room for improvment.	2018-02-06 18:29:25
686	Don't know	2018-02-06 18:34:46

687	a good tool	2018-02-06 18:25:13
688	I couldn't go back to living without it.	2018-02-06 18:25:45
689	Useful tool	2018-02-06 18:26:32
690	Like an extra pair of hands, useful for multitasking or speeding up the ability to complete a task that I would otherwise be doing myself	2018-02-06 18:25:24
691	invaluable	2018-02-06 18:30:02
692	Google Home is very useful, but has issues picking up your voice when there is music playing or a lot of ambient noise	2018-02-06 18:22:14
693	A pretty cool tech toy	2018-02-06 18:22:41
694	Useful for controlling my lights	2018-02-06 18:25:09
695	Very helpful, yet needs more improvement	2018-02-06 18:19:21
696	Life Changing	2018-02-06 18:18:55
697	Useful	2018-02-06 18:17:37
698	It's an easy way to get answers & control the house	2018-02-06 18:20:02
699	useful, kitchen helper	2018-02-06 18:19:30
700	Fun	2018-02-06 18:17:02
701	Bought it for the novelty. Surprised at how useful it has been.	2018-02-06 18:16:03
702	Like a small child, most of the time okay, buy sometimes I wanna punt it.	2018-02-06 18:15:18
703	Handy to have not essential.	2018-02-06 18:15:17
704	Google Home	2018-02-06 18:14:18
705	Useful toy	2018-02-06 18:13:27
706	Helpful. Not necessary.	2018-02-06 18:12:25
707	With the current state of machine learning and algorithms, I classify the smart speaker as a tool at this time.	2018-02-06 18:12:08
708	Fun you. Hopefully much more useful in the future	2018-02-06 18:12:48
709	Helpful, entertaining, useful. I occasionally can't figure out how to get it to do what I want it to do	2018-02-06 18:14:14
710	A great convenience when used as a part of a smart home	2018-02-06 18:14:22
711	It's a machine	2018-02-06 18:10:50

712	Perfekt for my needs	2018-02-06 18:10:51
713	It's just an interface to services on the 'net.	2018-02-06 18:09:53
714	Very handy at times, but not essential. Lots of room for further improvements	2018-02-06 18:11:49
715	It's great but could be better (and will get better in time)	2018-02-06 18:10:22
716	Essential to modern homes	2018-02-06 18:10:15
717	Useful tool	2018-02-06 18:08:28
718	Useful but at the same time annoying	2018-02-06 18:09:23
719	Still needs work. Answers most simple questions but is far from optimal.	2018-02-06 18:09:23
720	Its a box that does stuff	2018-02-06 18:05:42
721	Good at what it does, but not much else	2018-02-06 18:04:05
722	My GH is cutting edge, a beautiful experiment.	2018-02-06 18:04:53
723	Like others have said I wish there was a way to turn off the confirmation on most things. I do enjoy walking in to my house and listening to news.	2018-02-06 18:07:13
724	Not very noticeable, which is good. Should be there to help, but not something I want contantly being used	2018-02-06 18:03:23
725	A very useful machine. It has such potential that is unseen still. They say you can 'teach it' but I haven't seen it learn anything yet. It does its job... but nothing more.	2018-02-06 18:03:49
726	It's a fun party trick, a good timer, but lacks the intelligence to be truly useful.	2018-02-06 18:05:11
727	Useful	2018-02-06 18:02:54
728	Convenient, but not necessary	2018-02-06 18:03:16
729	After two years of use it has become an integral part of my daily routine.	2018-02-06 18:03:51
730	Useful	2018-02-06 18:02:21
731	Mrs. Slave	2018-02-06 18:01:16
732	I love that it will play and broadcast music. I also love using it to control the lights in my home.	2018-02-06 18:01:14
733	Good	2018-02-06 18:00:51
734	Utilitarian, in a good way	2018-02-06 17:58:11
735	Beautiful. Can't wait to see what Google had in store for the future of AI	2018-02-06 17:58:57

736	More useful than expected but i'm not dependant, it's proven to add to the experience with my current devices rather than replace them	2018-02-06 18:00:44
737	Works well, but still have to form questions in specific ways to get it to do what you want.	2018-02-06 17:57:06
738	Good but needs improvement	2018-02-06 18:04:50
739	Smart enough to get some things done well and dumb enough to mess up some really easy things.	2018-02-06 17:59:06
740	Good	2018-02-06 17:57:40
741	Really smart half the time, really dumb the other half	2018-02-06 17:57:43
742	A robot that helps me with tasks such as controlling lights, reading news, and informing me about weather and my daily tasks.	2018-02-06 17:57:35
743	A machine that was created to do my bidding	2018-02-06 17:55:38
744	Useful	2018-02-06 17:54:07
745	Something I didn't need but now don't want to be without.	2018-02-06 17:55:43
746	Handy	2018-02-06 17:53:25
747	a prototype for future technology	2018-02-06 17:51:18
748	Useful, though not particularly intelligent.	2018-02-06 17:51:10
749	A welcome addition but still needs work	2018-02-06 18:06:51
750	a neat toy	2018-02-06 17:50:17
751	Nice to have but not a must for me	2018-02-06 17:50:04
752	A gimmick that became indispensable.	2018-02-06 17:50:09
753	Looks like a fancy air freshener	2018-02-06 17:46:49
754	Frustratingly literal	2018-02-06 17:46:52
755	information and music on demand, smart home controller	2018-02-06 17:46:04
756	Helpful media player	2018-02-06 17:44:19
757	I really like that google assistant has been getting smarter and easier to use. They keep adding features such as broadcast that we use all the time. We look forward to it advancing even more!	2018-02-06 18:13:21
758	Don't know	2018-02-06 17:41:51
759	Makes day to day tasks easier	2018-02-06 17:40:47

760	Good assistant	2018-02-06 17:41:57
761	pretty cool	2018-02-06 17:40:35
762	Useful when I need to use it	2018-02-06 17:41:44
763	It listens for voices and responds.	2018-02-06 17:39:30
764	Usefull hands free device to get lights, play music, set alarms, order uber.	2018-02-06 17:40:20
765	Useful to my daily needs	2018-02-06 17:39:24
766	I have a hard time finding things to use it for besides music. I'd probably be better off with a chromecast and speakers.	2018-02-06 17:39:48
767	Useful for converging smarthome hardware and services.	2018-02-06 17:38:18
768	Not so smart	2018-02-06 17:38:34
769	Handy to have, young kids love to ask questions even if it won't give responses but has a hard time hearing their quieter voices.	2018-02-06 17:37:44
770	I don't know	2018-02-06 17:36:35
771	Has lots of potential	2018-02-06 17:37:28
772	Useful, Ease of Use, Nifty	2018-02-06 17:37:09
773	It is a tool.	2018-02-06 17:36:27
774	It's a great way to control a smart home	2018-02-06 17:36:56
775	Almost there	2018-02-06 17:37:27
776	Integration tool for Smart Home set up and on demand music. Not very responsive to me so normally use text commands to ensure efficiency	2018-02-06 17:35:58
777	Incredibly useful tool that has become a very prominent part of my daily routine.	2018-02-06 17:06:21
778	useful when associated with other connected devices like the chromecast or philipps hue. If not, it's just a wireless speaker.	2018-02-06 17:04:58
779	A useful tool with effective use of a niche interface	2018-02-06 17:05:00
780	Like a personal assistant, except you hired them from care in the community.	2018-02-06 17:03:06
781	Helpful	2018-02-06 17:03:21
782	Google is by far the best on the market	2018-02-06 17:02:03
783	Good for my family's needs, works well.	2018-02-06 17:03:25

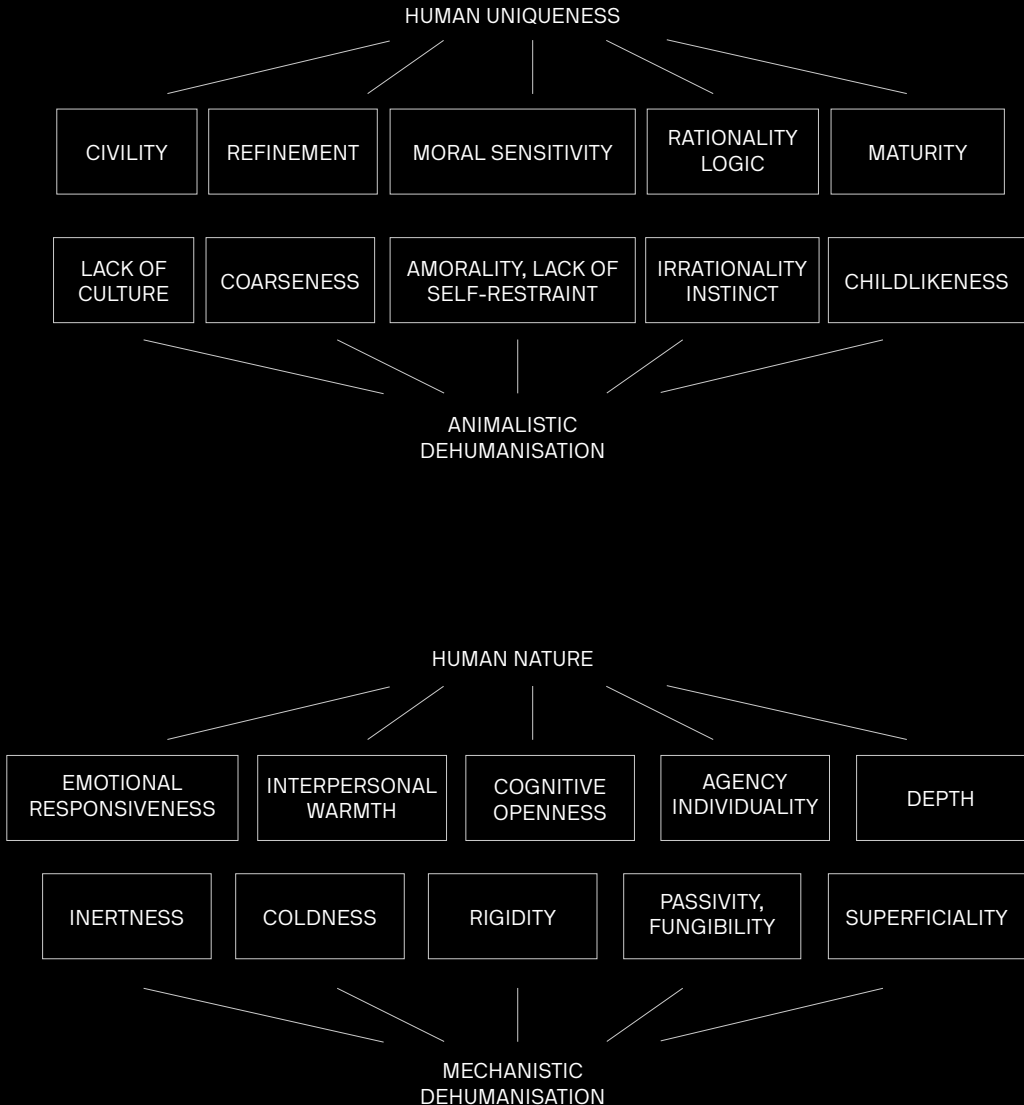
784	Great	2018-02-06 17:02:18
785	A toy unless you invest in more hardware to control your home	2018-02-06 17:02:17
786	its worth the money	2018-02-06 17:02:41
787	Convenient	2018-02-06 17:02:37
788	Pretty neat	2018-02-06 17:00:31
789	A useful device	2018-02-06 17:02:16
790	Very helpful but not completely necessary.	2018-02-06 17:00:47
791	awesome	2018-02-06 17:00:24
792	I hope they preserve my privacy.	2018-02-06 17:04:26
793	Input device for various small repetitive tasks	2018-02-06 17:02:11
794	Quick, friendly, knowledgeable	2018-02-06 17:01:45
795	the future is awesome	2018-02-06 16:59:50
796	Helpful, funny, and a learning tool	2018-02-06 17:04:59
797	It's a fancy toy. But it does some neat things that make life easier.	2018-02-06 16:59:11
798	Does almost everything I need it to do, and everything it hasn't been able to do has come in a software update	2018-02-06 16:58:02
799	It should do what I tell it and not ask questions.	2018-02-06 16:58:54
800	Nice gadget, but doesn't support my language (Danish) yet, so it's a bit frustrating to use.	2018-02-06 16:57:32
801	It's a work in progress. It generally helps me in the way I need, but sometimes will give random answers to straightforward questions, or can't complete my requests. It's made access to my music so easy that I haven't turned on my bluetooth audio since I got the smart speaker.	2018-02-06 16:57:03
802	Fun beyond just a novelty	2018-02-06 16:56:33
803	Stubborn	2018-02-06 16:54:54
804	Very useful	2018-02-06 16:55:09
805	The most basic yet useful home automation solution.	2018-02-06 16:53:47
806	Just right	2018-02-06 16:54:06
807	Ready to serve	2018-02-06 16:52:47

808	Not entirely necessary, but incredibly useful	2018-02-06 16:54:12
809	Useful	2018-02-06 16:52:01
810	Very effective tool	2018-02-06 16:52:06
811	h h	2018-02-06 16:49:08
812	Big promises that fall short of expectations	2018-02-06 16:48:50
813	Like a nice friend who does things you ask them to do	2018-02-06 16:47:34
814	A very useful tool which still needs a lot of work to perfect.	2018-02-06 16:45:07
815	Best thing ever	2018-02-06 16:43:29
816	A useful interface for accessing and manipulating digital assets	2018-02-06 16:43:20
817	Sometimes misunderstands simple requests	2018-02-06 16:45:13
818	Very cool	2018-02-06 16:42:38
819	Good for what it does	2018-02-06 16:42:07
820	A very useful tool that humans can use to delegate smaller responsibilities to so that the human can focus on big picture things	2018-02-06 16:40:44
821	Helpful	2018-02-06 16:38:56
822	The speaker is helpful and useful for general information and smart home control.	2018-02-06 16:40:39
823	Convenient, but still growing.	2018-02-06 16:38:46
824	Very helpful	2018-02-06 16:39:03
825	Automation and voice commands	2018-02-06 16:39:12
826	Useful addition to my house	2018-02-06 16:38:41
827	It is a series of programmed responses in a box	2018-02-06 16:37:45
828	Frustrating, but I knew that going in as an early adopter.	2018-02-06 16:35:43
829	It's helpful. Not a necessity but is something that makes annoying little tasks less so.	2018-02-06 16:38:42
830	Useful gadget	2018-02-06 16:35:27
831	Fun, useful, easy to use	2018-02-06 16:41:16
832	An unessential tool made essential.	2018-02-06 16:33:20
833	potentially lifechanging	2018-02-06 16:31:31

834	Rigid, barebones, immature.	2018-02-06 16:31:20
835	A one way conversation	2018-02-06 16:31:56
836	The Google Assistant is a good piece of programming that makes interaction with devices more fun because it's "human-ish".	2018-02-06 16:32:38
837	It makes life easier. :)	2018-02-06 16:29:41
838	It's a hands free/eyes free UI for entertainment and simple productivity tasks.	2018-02-06 16:27:13
839	It funtions, your last question... Why are you trying to get people to attrubute emotions to an algorithm?	2018-02-06 16:29:05
840	An interesting emerging/developing technology.	2018-02-06 16:26:01
841	Honestly the most useful thing I've bought in a while. I use it to control all my lights and even have it set so that it turns off my lights when I leave the house.	2018-02-06 16:24:49
842	Up and coming	2018-02-06 16:23:04
843	A handy tool	2018-02-06 16:21:50
844	Useful	2018-02-06 16:20:08
845	Can't wait to see the technology grow.	2018-02-06 16:21:23
846	Intuitive and helpful	2018-02-06 16:18:16
847	Helpful, but you need to speak it's language patterns as it hasn't learned ours	2018-02-06 16:18:36
848	It is not what I would consider "AI" yet. It only knows what it can search the internet for - not understand what it means	2018-02-06 16:18:12
849	Useful, amusing, lots of potential	2018-02-06 16:20:10
850	As useful as I make it	2018-02-06 16:16:28
851	Almost there.	2018-02-06 16:16:55
852	Sometime helpful other times not. It's a good start	2018-02-06 16:15:49
853	Cool technology that has a long way to go.	2018-02-06 16:14:31
854	Useful	2018-02-06 16:12:35
855	Does it's job.	2018-02-06 16:08:53
856	A fun, but not necessary, addition to the home.	2018-02-06 16:08:46
857	Great purchase when coupled with premium music services and smart home items.	2018-02-06 16:10:04
858	Makes my life easier	2018-02-06 16:06:04

859	Software platform has to mature further to get better use.	2018-02-06 16:34:21
860	Very nice to have, but not life-changing. I could live without it if I needed to.	2018-02-06 16:01:10
861	It is a useful device for voice control and basic information	2018-02-06 15:57:25
862	A welcome addition to the home.	2018-02-06 15:56:31
863	Great virtual assistant	2018-02-06 15:50:51
864	works great	2018-02-06 15:45:32
865	A fun toy but not a member of my family!	2018-02-06 15:17:08
866	Handy, but limited.	2018-02-06 15:07:52
867	Life changing	2018-02-06 15:03:25
868	It's a useful gadget that makes some of my tasks easier	2018-02-06 15:03:57
869	Sadly underutilized and prone to hearing her name more often than it is said.	2018-02-06 14:32:54
870	Not quite perfected tech	2018-02-06 14:23:58
871	Makes it easier to find information and programs	2018-02-06 14:07:19
872	Not quite good enough.	2018-02-06 13:56:37
873	She does the job.	2018-02-06 13:48:41
874	Annoying semantic based pattern recognition device. Usually fails after 3 simple words. Semantic engine behind it is very primitive. Voice recognition is fair but too many apps waiting behind it so wrong assumptions and semantic fails are very common. It requires specific world order (e.g.: "play trance" and "play trance music" is not interpreted on the same way). After 2-3 weeks a started to use/direct it with my phone as I started to become boring listen the same music channels what it can understand. Usually it starts to play the same amazon music channel when it fails to recognise (saying: "Here is a music channel what you may like."). It is just a wireless speaker now with an annoying fancy feature what I may prefer to switch off. Despite the time and effort to learn and set-up the device and applications behind it.	2018-02-06 12:58:06

E. HUMAN NATURE AND HUMAN UNIQUENESS SCALE



Proposed links between conceptions of humanness and corresponding forms of dehumanization (Haslam, 2006, p. 257)

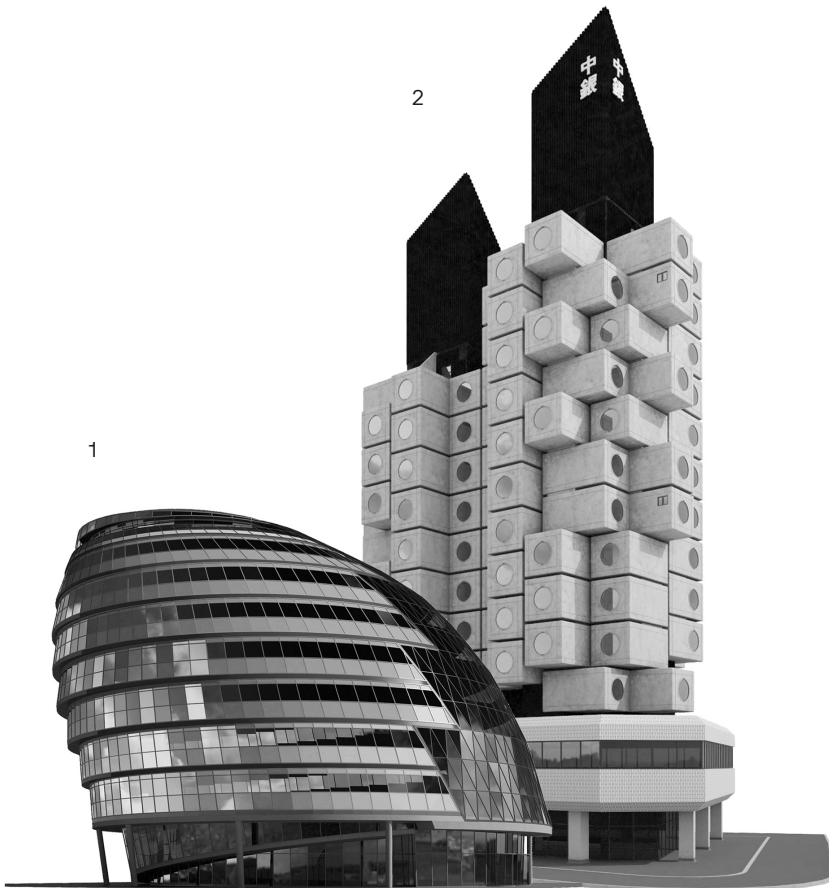
	ANIMALISTIC	MECHANISTIC
Form of Denied Humanness	Uniquely Human	Human Nature
Implicit Contrast	Animals	Automata
Prototypical Domains	Interethnic relations, disability	Technology, biomedicine
Exemplary Theories	Infra-humanization	Value based, objectification
Emotion	Disgust, contempt	Disregard, indifference
Semiotics	Vertical Comparison	Horizontal comparison
Essentialism	Nature of difference between perceiver and target	Content of attributed difference between perceiver and target
Social Context	Primarily intergroup	Intergroup and interpersonal
Relational Definition	Communal sharing	Asocial
Cognitive Modality	Natural history / folk biology	Technical
Behavioural Explanation	Desire based	Cause or causal history based

F. 'JUST LIKE ME' – ACTIVITY

1 Landmark 01:
London Town Hall
(UK)

2 Landmark 02:
Nakagin Capsule
Tower (Japan)

05





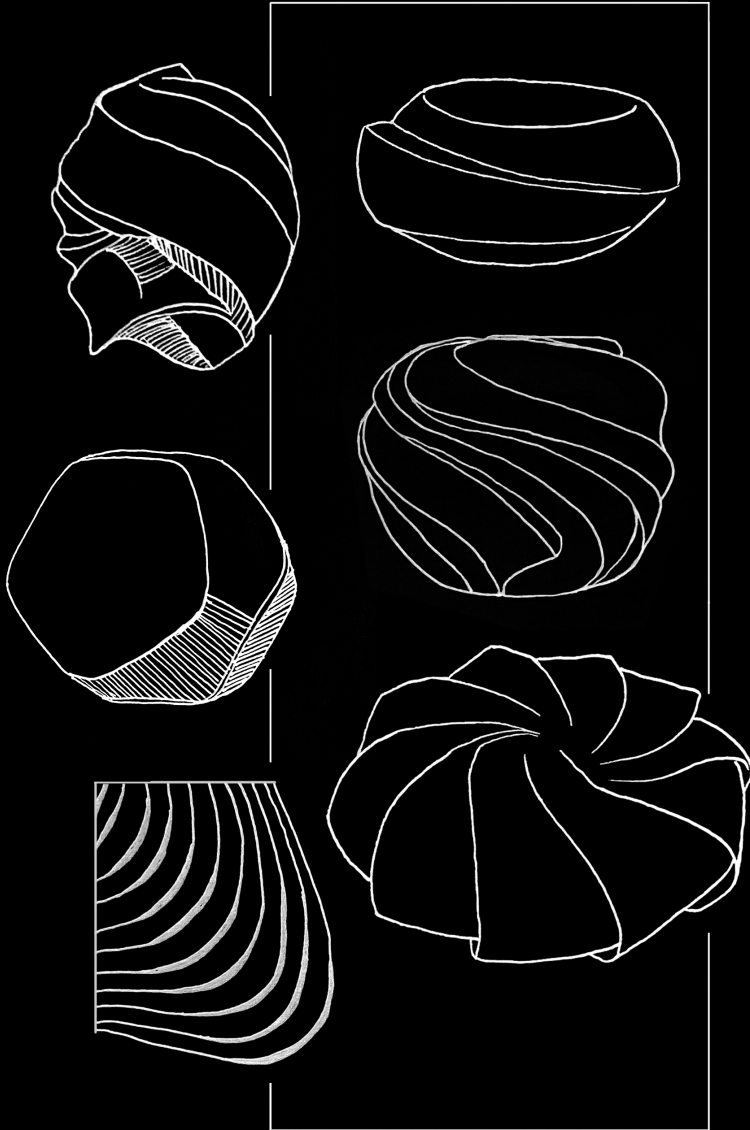
5 Landmark 05:
Oriental Pearl Tower
(China)

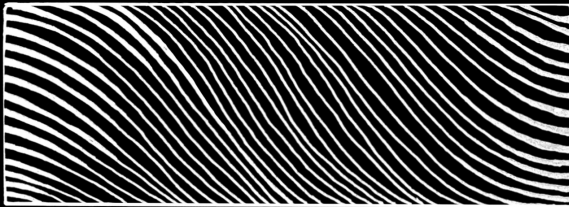
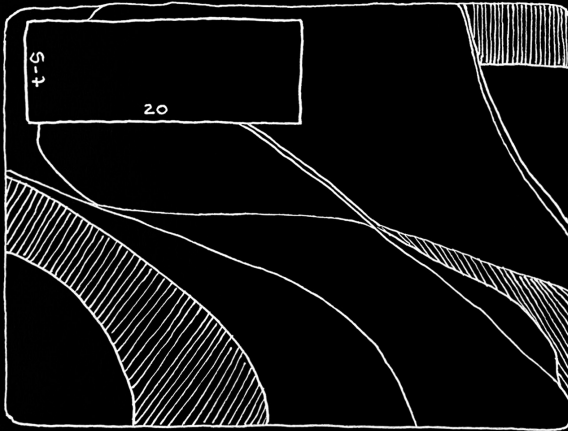
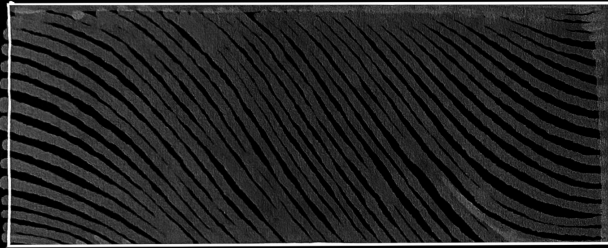
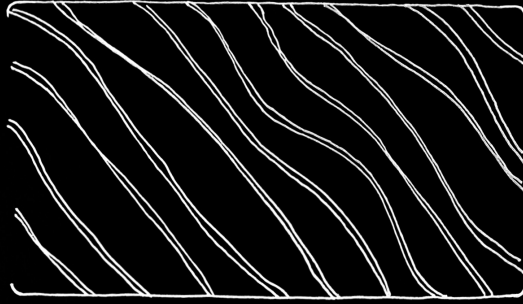
6 Landmark 06:
Lotus Temple
(India)

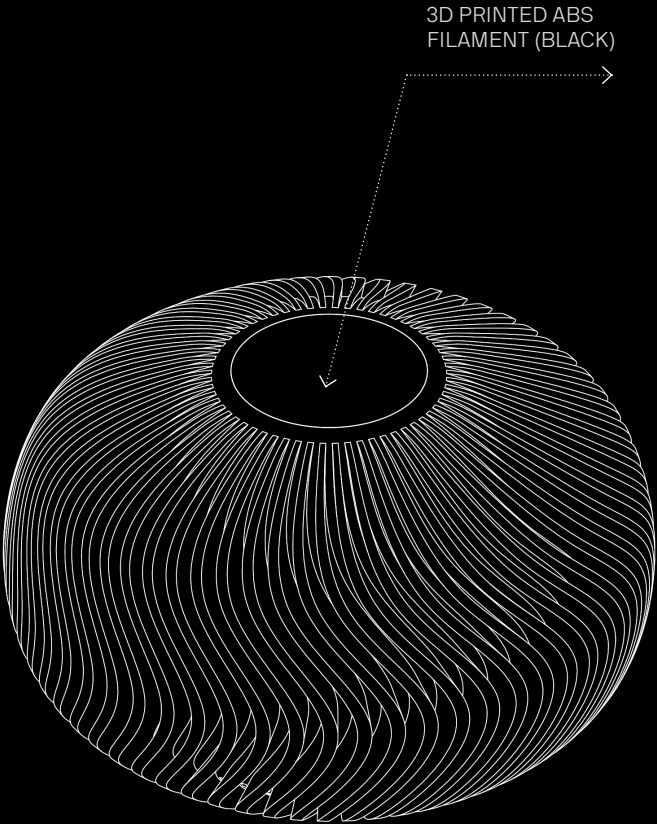
05

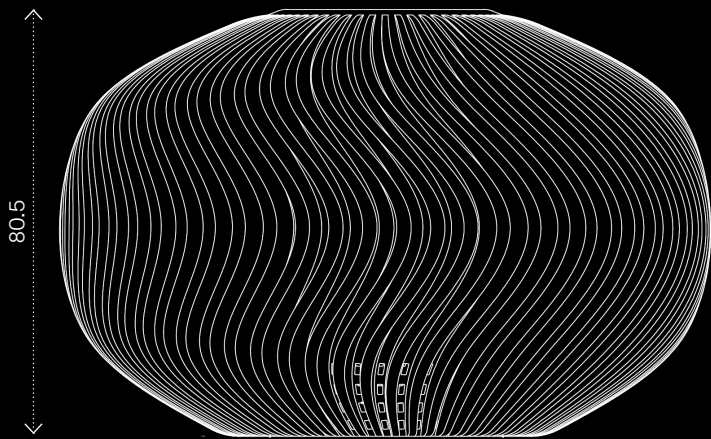
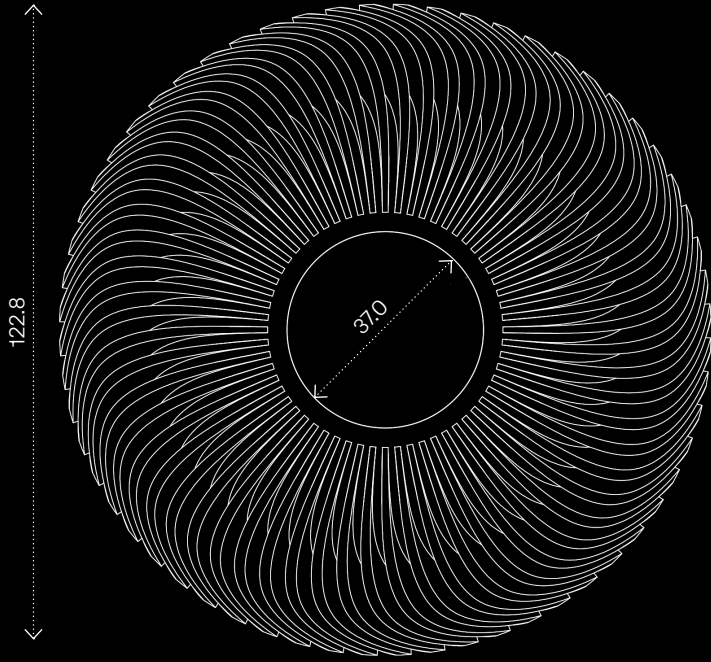


G. 'JUST LIKE ME' – DEVICE DESIGN



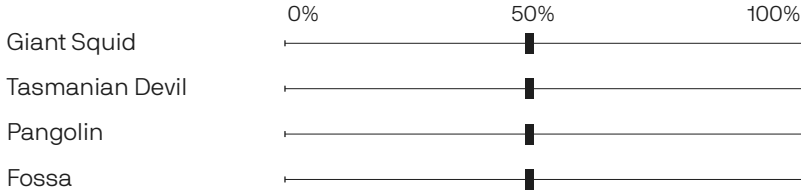






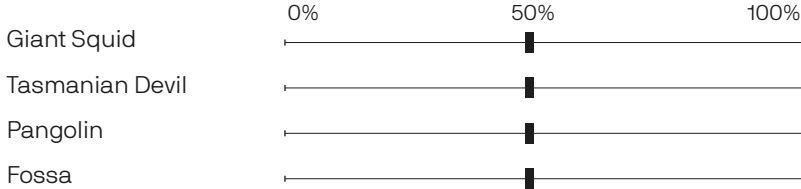
H. 'BUT NOT EXACTLY...' SURVEY

Q1. Please estimate the AI's likelihood (in %) of identifying the following animals:



This is a hypothetical question:

Q2. Please estimate another person's likelihood (in %) of identifying the following animals:



Q3. Please rate on a scale of 1 (not at all) to 10 (a lot) how much did you enjoy the game:

1 2 3 4 5 6 7 8 9 10

I. 'BUT NOT EXACTLY...' SURVEY SUMMARY

A. LITERAL CONDITION

AI AGENT				
#	Easy	Medium	Hard	Expert
1	98%	82%	28%	7%
2	80%	70%	40%	10%
3	100%	90%	35%	5%
4	85%	70%	35%	15%
5	95%	85%	30%	20%
6	79%	65%	11%	0%
7	90%	80%	30%	20%
8	95%	90%	25%	7%
9	95%	80%	55%	15%
10	100%	90%	65%	25%
11	90%	85%	35%	15%
12	100%	90%	35%	20%
13	80%	70%	45%	10%
14	73%	62%	10%	1%
Average	90%	79.2%	34.2%	12.1%

HUMAN				
#	Easy	Medium	Hard	Expert
1	92%	94%	18%	2%
2	85%	75%	20%	1%
3	100%	85%	15%	5%
4	90%	75%	25%	10%
5	85%	85%	60%	10%

6	82%	75%	5%	0%
7	90%	80%	15%	5%
8	80%	65%	20%	3%
9	85%	63%	41%	10%
10	90%	75%	25%	35%
11	96%	76%	14%	9%
12	95%	55%	20%	15%
13	75%	40%	25%	10%
14	63%	77%	18%	6%
Average	86.3%	72.9%	22.9%	8.6%

B. ROBOTIC CONDITION

AI AGENT

#	Easy	Medium	Hard	Expert
1	99%	82%	49%	21%
2	92%	79%	58%	42%
3	95%	80%	50%	20%
4	87%	73%	32%	2%
5	100%	90%	65%	35%
6	95%	90%	45%	10%
7	100%	100%	70%	45%
8	90%	95%	25%	10%
9	100%	100%	70%	40%
10	100%	80%	35%	9%
11	94%	100%	39%	5%
12	89%	76%	8%	3%
13	100%	90%	30%	13%
14	100%	85%	43%	19%
15	95%	89%	43%	12%
Average	95.7%	87.3%	44.1%	19.1%

HUMAN

#	Easy	Medium	Hard	Expert
1	87%	90%	27%	13%
2	96%	65%	41%	34%
3	80%	60%	45%	15%
4	93%	67%	23%	0%
5	95%	80%	25%	5%
6	85%	80%	40%	5%
7	95%	90%	30%	25%
8	80%	65%	10%	5%
9	90%	70%	60%	25%
10	95%	70%	25%	2%
11	82%	95%	7%	0%
12	87%	67%	3%	1%
13	90%	80%	15%	3%
14	95%	65%	26%	7%
15	90%	69%	17%	4%
Average	89.3%	74.2%	26.3%	9.6

A. LITERAL CONDITION SATISFACTION

Basic Statistics

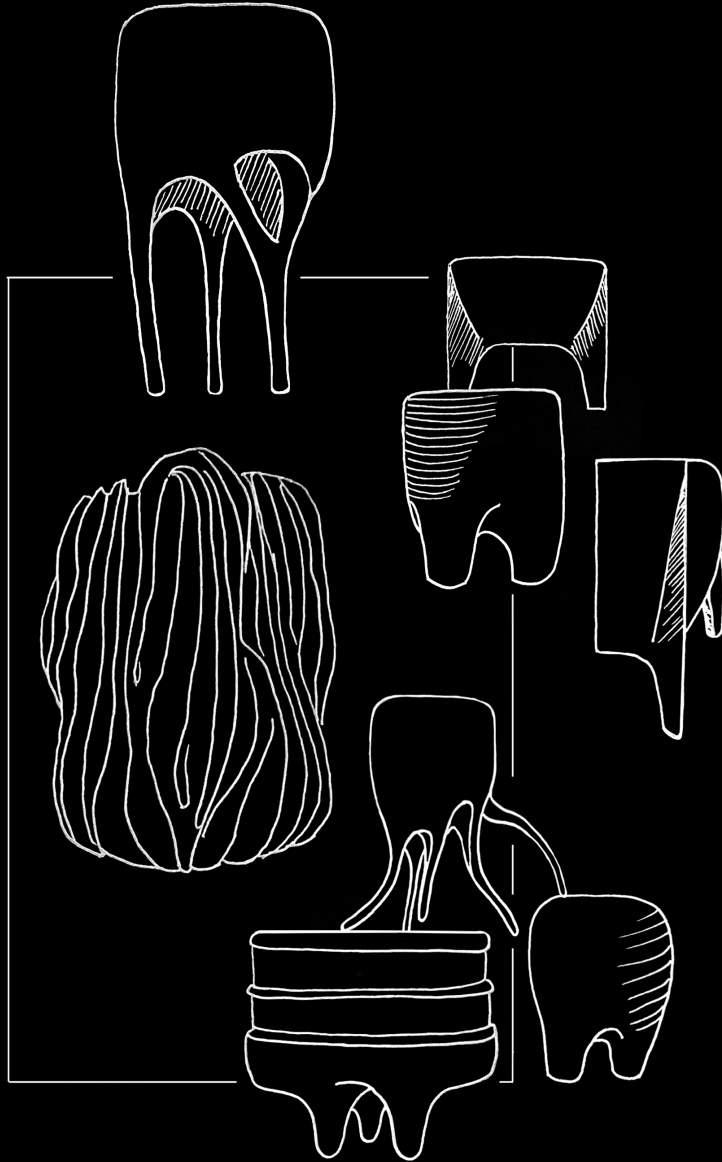
Minimum	Maximum	Median	Mean	Standard Deviation
7.00	10.00	9.00	9.21	0.89

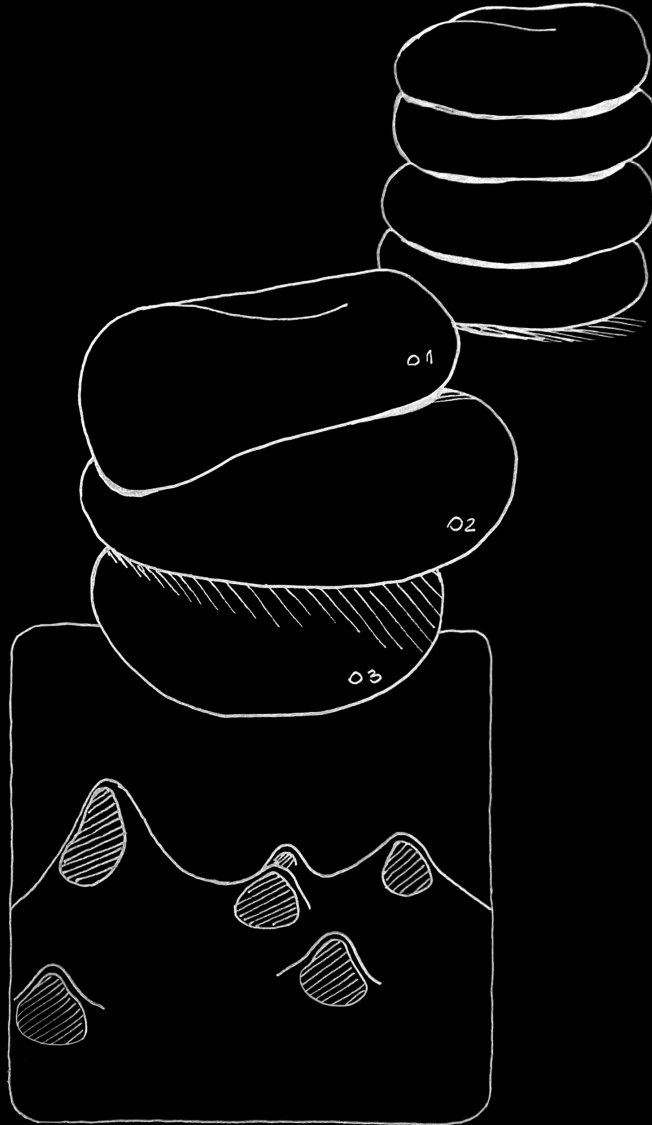
B. ROBOTIC CONDITION SATISFACTION

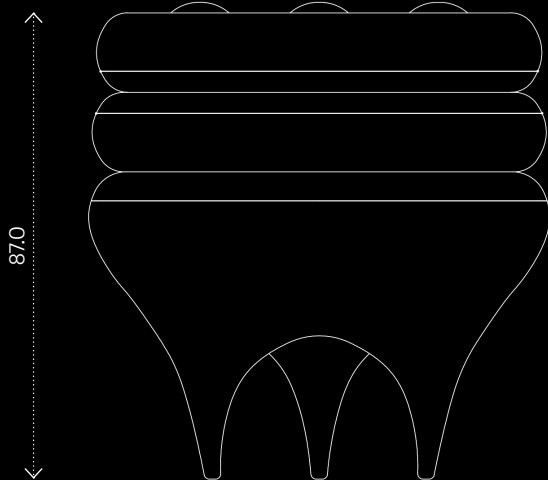
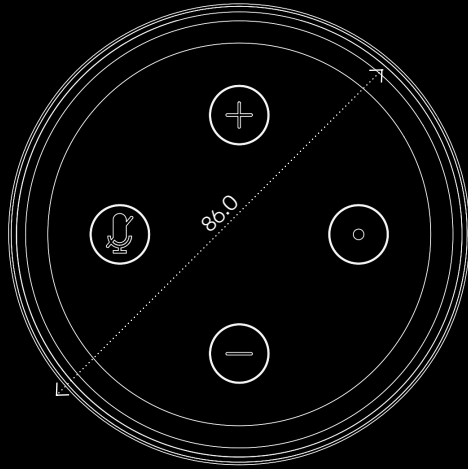
Basic Statistics

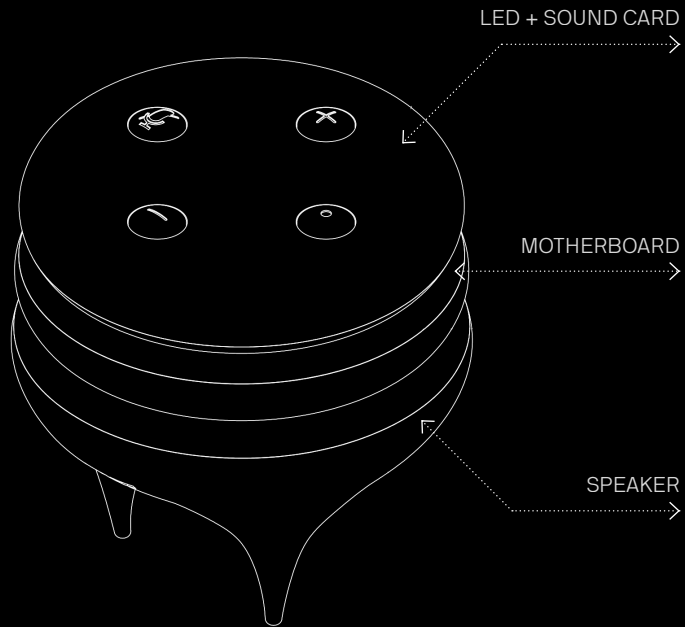
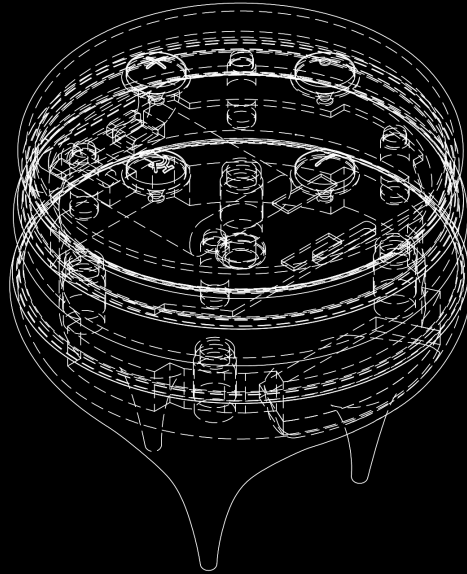
Minimum	Maximum	Median	Mean	Standard Deviation
3.00	10.00	7.00	6.93	1.91

J. 'BUT NOT EXACTLY...' DEVICE DESIGN









K. INTERSPACE: CONVERSATIONAL ANALYSIS

SIRI

Average Pitch: A [220Hz] Siri [210Hz]

		Duration (s)	Pause (s)	Pitch (Hz)	V (dB)	Breaks (%)
A	Hey Siri	0.8	1.0	241.6	70.3	9.708
	I want to see a movie	1.1	1.4	214.8	68.7	7.197
S	Here are some movies playing at theatres near you Bohemian Rhapsody A star is Born Halloween	5.8	0.8	204.6	66.1	31.315
	First Man Jonny English strikes again	2.4	1.1	203.2	65.4	35.462
A	Ca[n you sh..	0.3	0.3	272.8	70.0	0
S	Do] you want to hear the next 5 movies?	1.8	0.6	232.5	66.2	14.243
A	Can you show me the timetable of A star is born?	2.9	2.4	232.5	69.2	23.692
S	Here's A star is born	1.1	1.3	228.0	66.7	37.630
	Hard-drinking country music star Jackson Main discovers and falls in love with. A struggling but talented singer named Ally?	6.7	1.1	208.8	65.9	24.770
	Should I keep going?	0.9	1.3	240.6	63.6	5.506
A	When can I watch this movie?	2.1	1.6	224.1	68.1	18.879
S	Bohemian Rhapsody starts Rami Male[k Lu]cy Boyton and Gwilym Lee	4.4	1.2	198.0	66.4	26.301

A	*laughter*	0.7		245.2	63.7	16.045
	Hmmmm	0.8	1.2	165.5	54.6	0
	Maybe let's do it again	1.2	-	188.4	62.6	2.342

GOOGLE HOME

Average Pitch: J [140Hz] GH [215Hz]

		Duration (s)	Pause (s)	Pitch (Hz)	V (dB)	Breaks (%)
J	Err	0.6	0.7	103.3	72.2	0
	Hey Google	0.7	0.6	193.8	69.5	4.105
	What was the main song from Ferris Bueller's Day Off?	3.5	2.2	115.8	68.7	18.507
G	I searched the web and found this song Oh Yeah	2.5	0.8	211.2	67.2	19.831
J	*laughter*	0.6	1.0	418.3	58.1	0
	[laughter]Hey Google	1.1	0.7	282.3	68.8	11.009
	Can you play this song?	0.7	3.0	114.8	64.5	34.150
G	Sure. Here is some music on Spotify	2.2	0.4	223.3	67.0	23.219
J	Hey Google	0.6	1.2	124.1	69.8	18.102
	Stop	0.4	0.5	191.5	72.0	26.979
	That's not what I said.	1.0	1.4	151.3	66.9	14.017
G	Sorry about that. I'm still learning.	2.1	0.3	219.4	66.3	21.133

Average Pitch: J [140Hz] GH [215Hz]

		Duration (s)	Pause (s)	Pitch (Hz)	V (dB)	Breaks (%)
	[*plays music*					
J	OK. Hey.	0.7	0.4	105.9	63.8	37.432
	Hey Google]	0.6	0.8	129.2	70.8	0
	STOP	0.4	3.8	239.9	72.3	24.159
	Hey Google	0.5	0.7	115.7	68.1	6.075
	Play the main song from Ferris Bueller's Day Off	2.2	3.2	112.9	67.1	13.146
G	Alright, here is Quintet in E Major, Opus 11, number 5, G275 to 3	6.0	0.7	215.8	67.4	17.036
	Minuetto, Ferris Bueller's Day Off on Spotify	3.0	2.4	215.3	67.3	24.991
	plays [music					
J	Oh yeah, this was in Ferris Bueller's Day Off	2.5	1.4	168.3	63.3	20.434
	Huh. Not the one I was thinking about	1.9	0.7	107.7	57.7	60.340
	Err. Hey Google]	0.9	0.9	142.1	72.7	7.209
	STOP. That's good. Thank you.	1.3		116.7	64.4	35.517

ALEXA

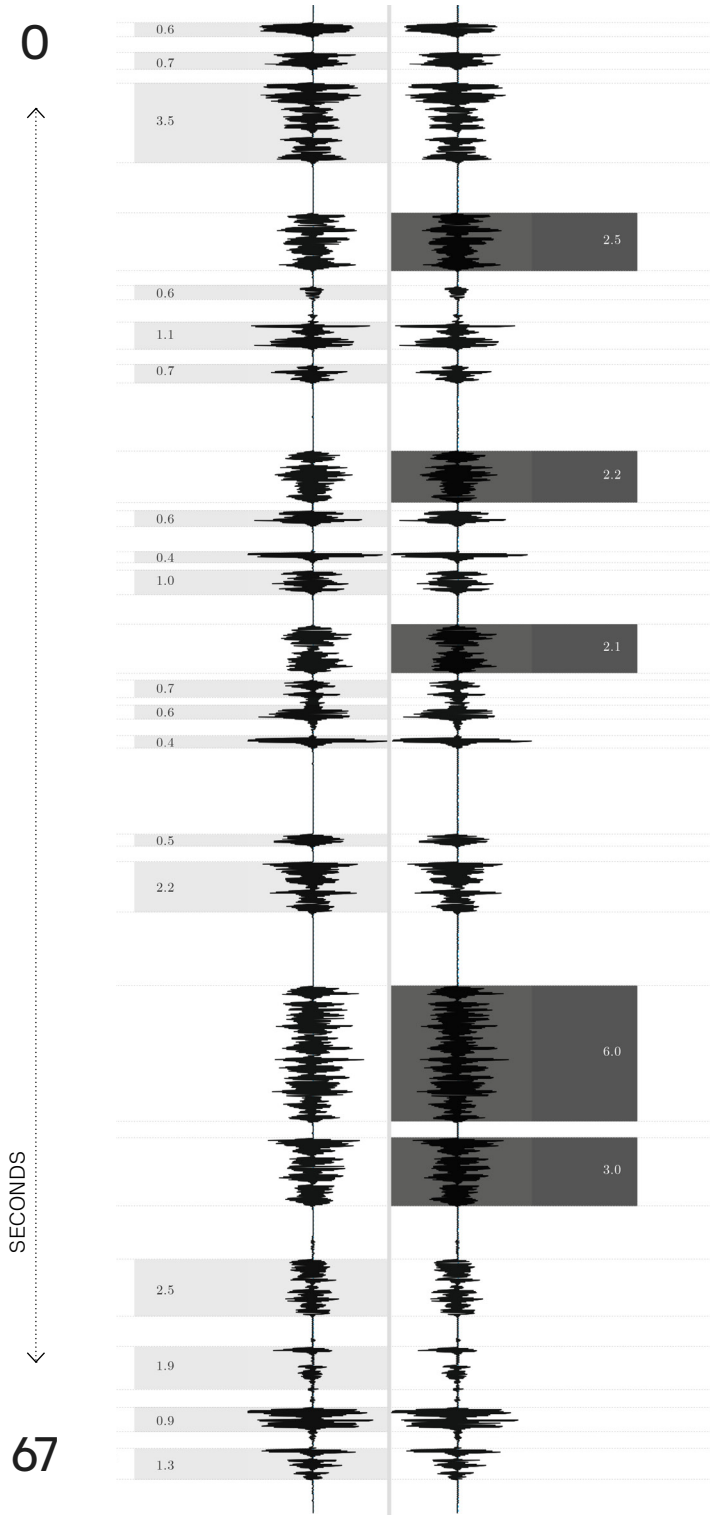
Average Pitch: A [220Hz] Siri [210Hz]

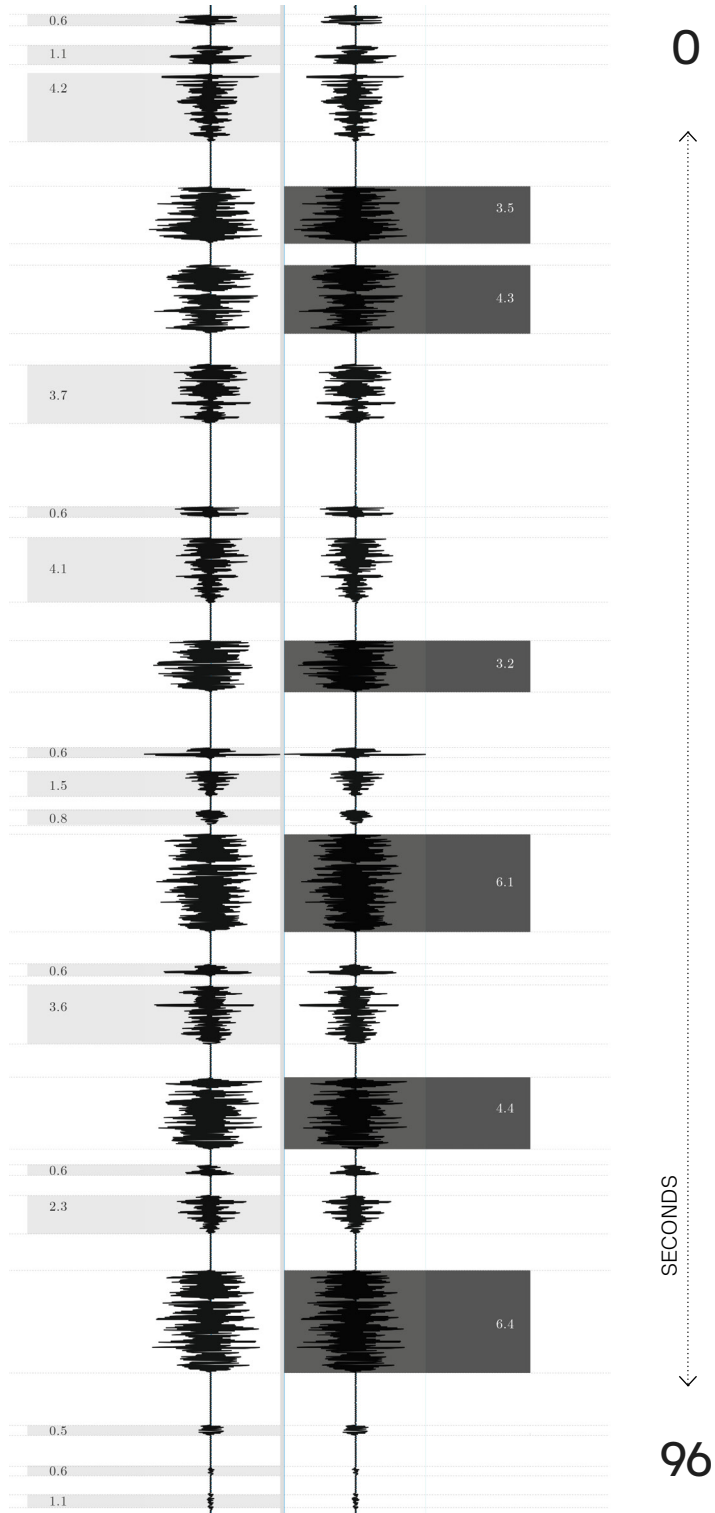
		Duration (s)	Pause (s)	Pitch (Hz)	V (dB)	Breaks (%)
R	Alexa	0.6	1.3	224.0	65.0	26.789
	Can you please tell me	1.1	0.6	305.4	66.1	12.160
	what, at what cinema nearby I could watch Mary Queen of Scots?	4.2	3.0	220.4	63.3	25.145
A	Mary Queen of Scots is playing at one theatre nearby today	3.5	1.5	183.7	69.9	19.261
	At Cineworld Leicester Square, there is one screening at 8:50pm	4.3	2.0	191.0	68.7	27.789
R	Is it playing anywhere tomorrow that is closer to my home?	3.7	5.4	235.2	65.3	14.804
	Alexa??	0.6	1.4	259.0	65.6	15.590
	Is it playing anywhere tomorrow that is closer to my current location?	4.1	2.5	228.7	63.7	22.330
A	I can't find cinemas near you playing "at anywhere" tomorrow	3.2	3.7	178.3	70.3	17.638
R	Alexa??	0.6	0.9	251.5	68.2	11.315
	Is it playing at Rich Mix	1.5	1.0	249.8	61.3	13.474
	Anywhere soon	0.8	0.7	184.9	59.1	15.685
A	Based on the current cinema schedule, I couldn't find screening for "anywhere with you" at Rich Mix cinema tomorrow	6.1	2.2	186.5	69.0	28.233
R	Alexa	0.6	0.7	242.5	65.4	24.321

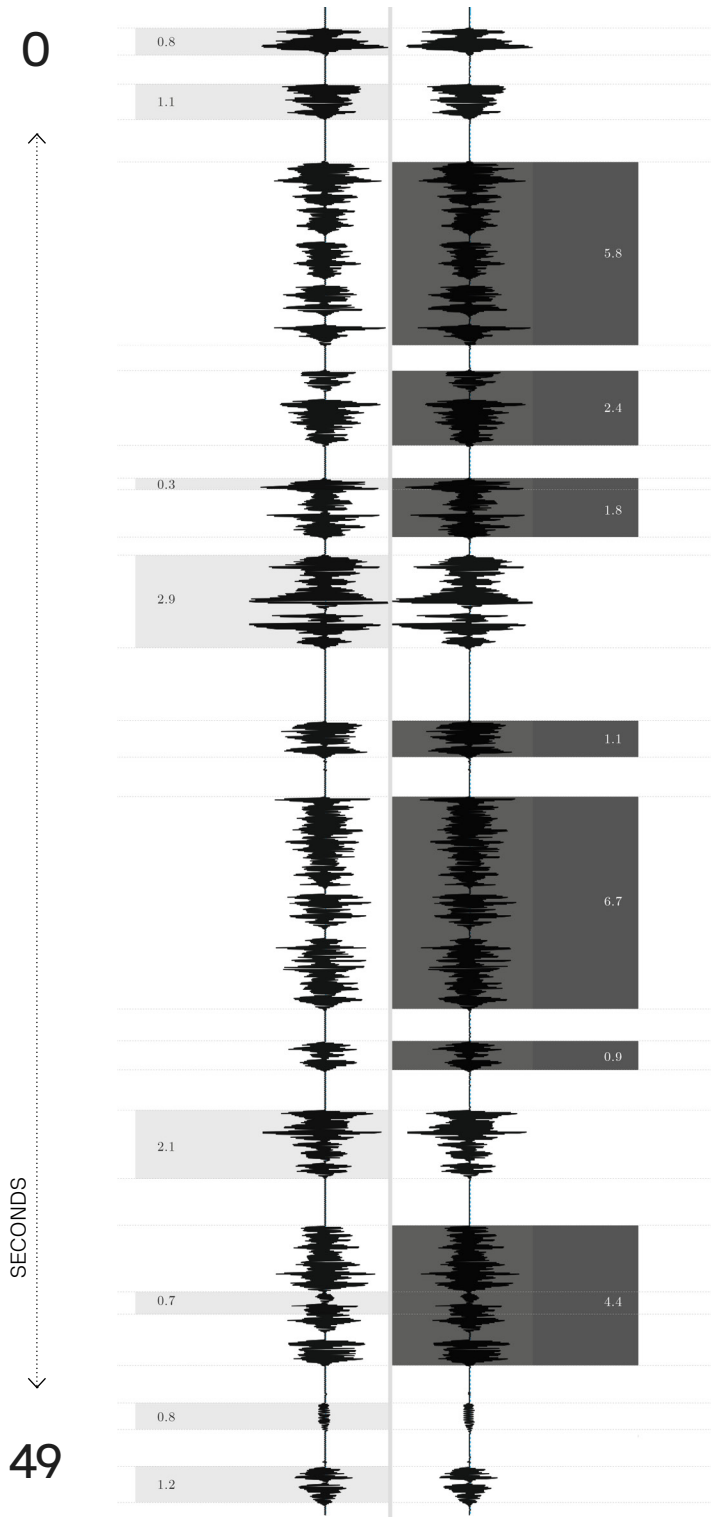
Average Pitch: A [220Hz] Siri [210Hz]

		Duration (s)	Pause (s)	Pitch (Hz)	V (dB)	Breaks (%)
	Is Mary Queen of Scots playing at Rich Mix anytime soon?	3.6	2.2	209.8	63.2	34.173
A	Sorry I couldn't find any cinema name "Rich Mix Anytime Soon" near you	4.4	1.1	185.6	69.6	20.141
R	Alexa?	0.6	1.3	220.8	61.5	26.199
	Is Mary Queen of Scots playing at Rich Mix.	2.3	2.4	223.7	62.7	36.258
A	Based on the current cinema schedule, I couldn't find screening for Mary Queen of Scots at Rich Mix cinema tomorrow	6.4	3.4	189.1	69.0	28.447
R	Okaaaay	0.5	2.1	245.0	59.8	0
	Yeah	0.6	1.0	187.9	44.5	12.822
	This wasn't very efficient	1.1	-	202.0	44.4	59.070

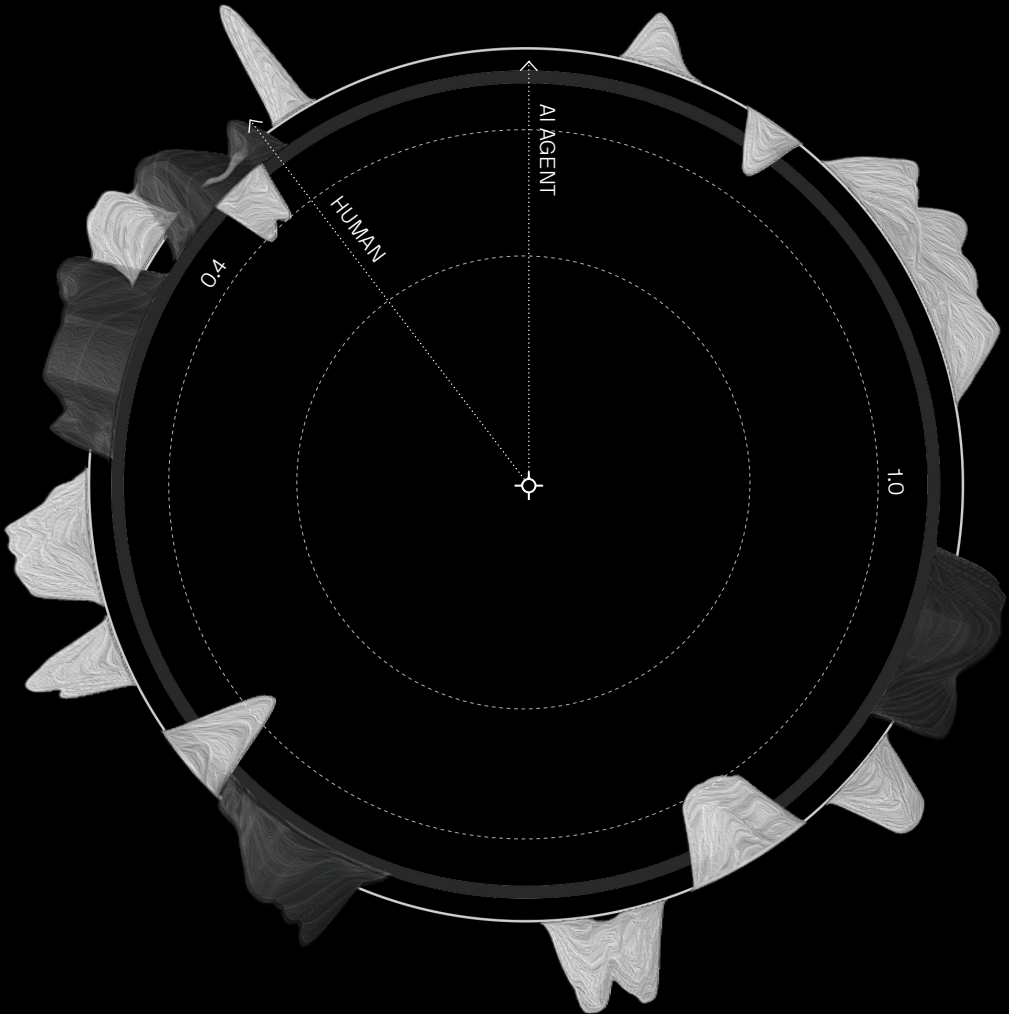
L. INTERSPACE: DATA ANALYSIS

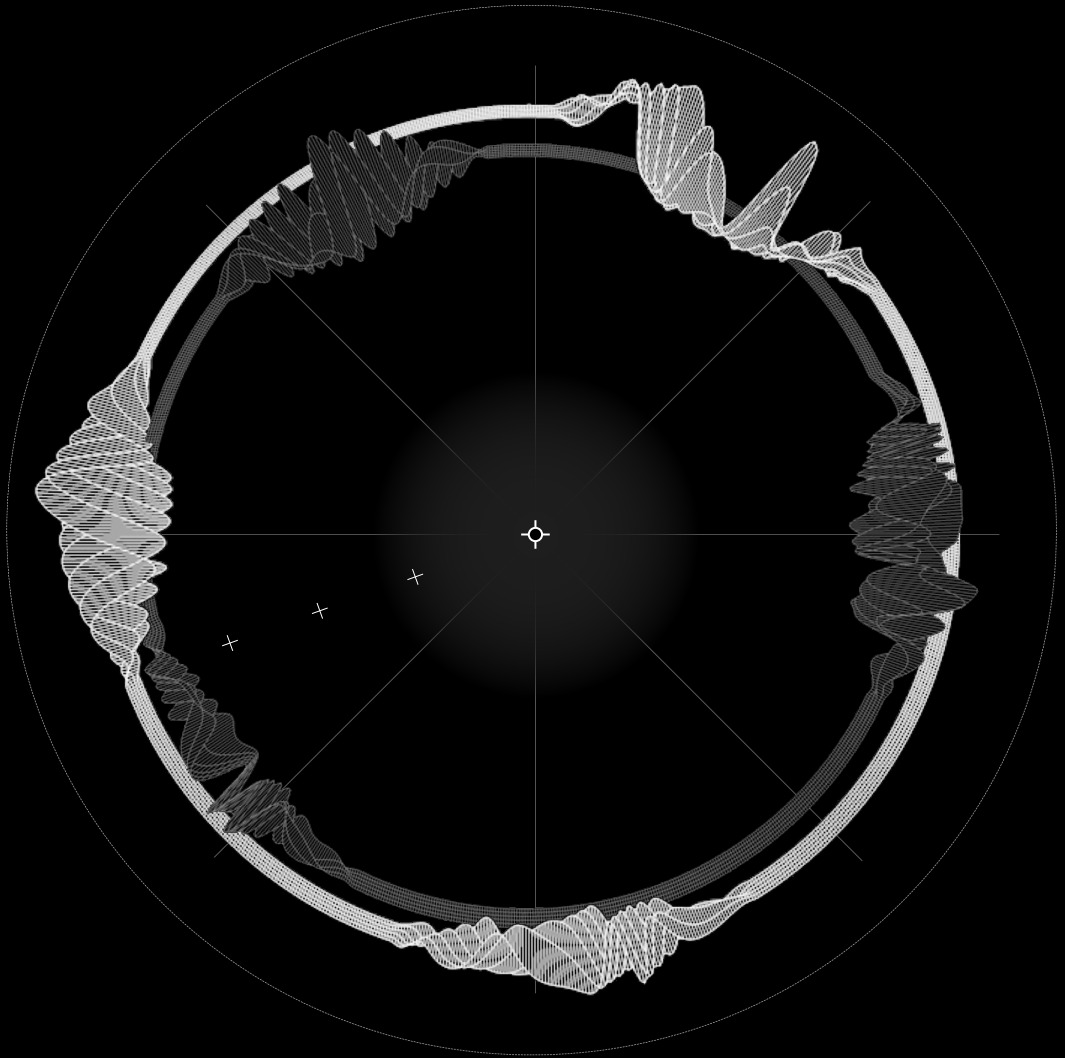


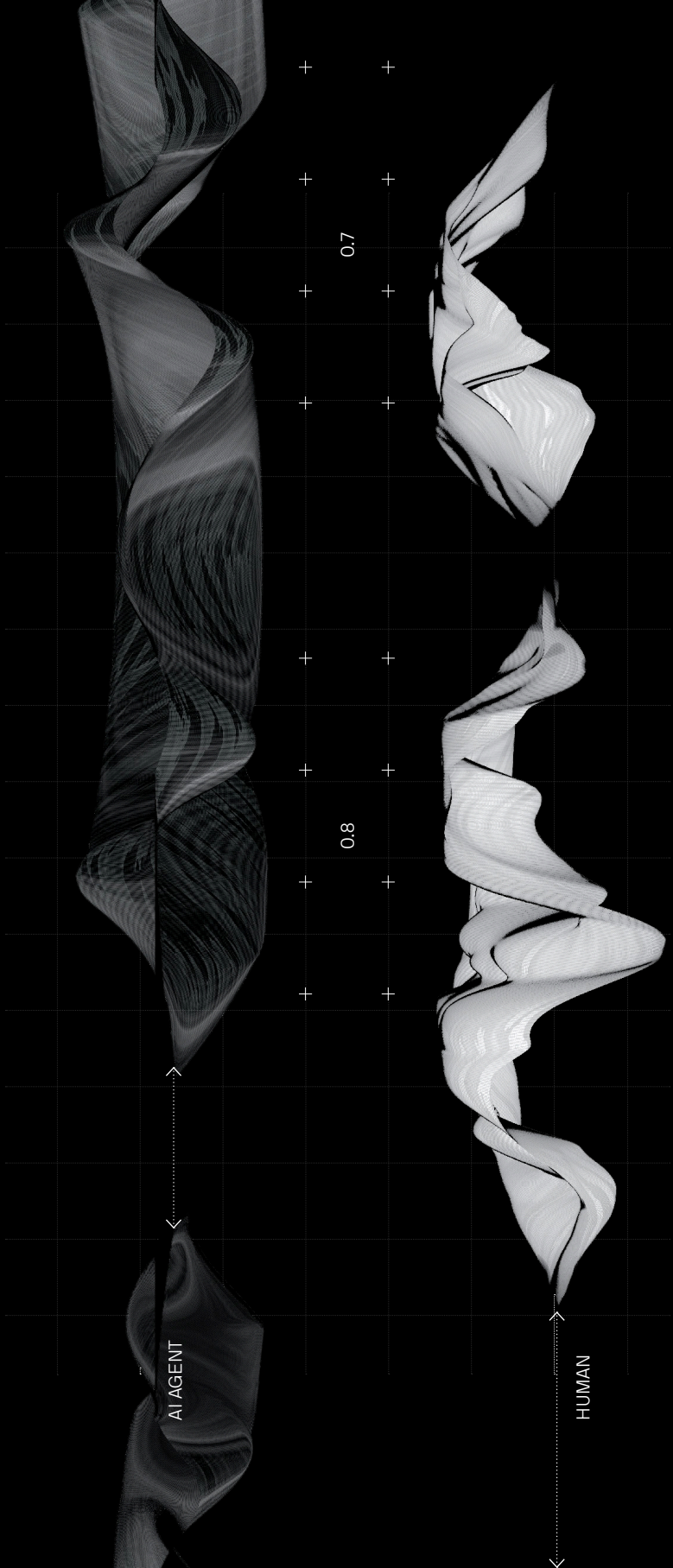


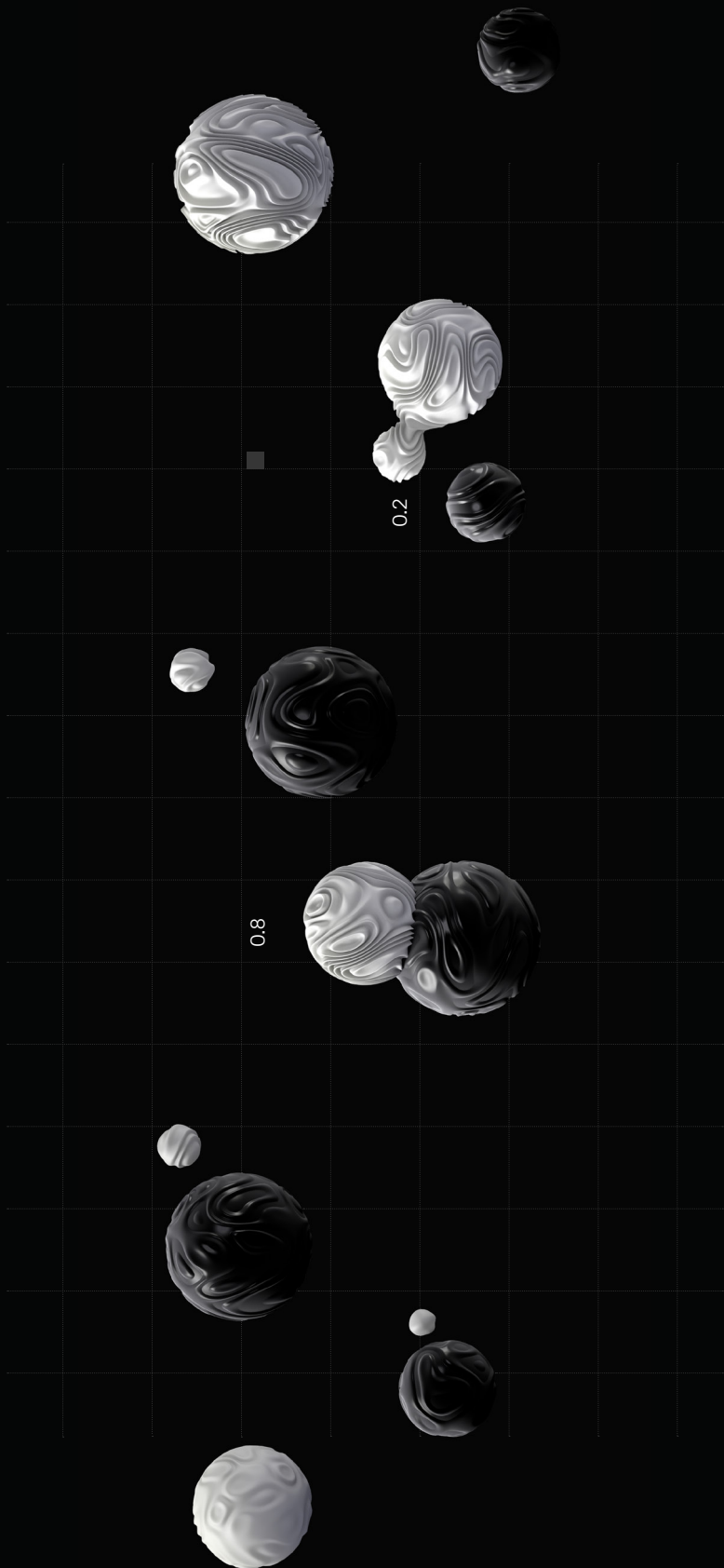


M. INTERSPACE: CONCEPT STUDIES









N. CONSENT FORM SAMPLE

PARTICIPANT PROJECT INFORMATION & CONSENT FORM

(One signed copy of this form should be retained by the Participant and one copy by the Project Researcher)

Date:

Dear Potential Participant,

I am a research student at the Royal College of Art and as part of my studies, I am conducting a study entitled, which is externally sponsored by Microsoft Research Cambridge. You are invited to take part in this research project which explores the nature of exchanges with social AI agents.

If you consent to participate, this will involve:

- + Naturally conversing with your virtual assistant. As you will aim to complete everyday tasks (such as asking for directions to the nearest shop, making lists or asking trivia questions), this interaction will be audio recorded with your consent.

Participation is entirely voluntary. You can withdraw at any time up to the point of publication and there will be no disadvantage if you decide not to complete the study. All information collected will be confidential. All information gathered will be stored securely and once the information has been analysed all individual information will be destroyed. Further, at no time will any individual be identified in any reports resulting from this study.

If you have any concerns or would like to know the outcome of this project, please contact my supervisor (Ashley Hall) at the following address: ashley.hall@rca.ac.uk

Thank you for your interest.

I (please print) have read the information above and all queries have been answered to my satisfaction. I agree to voluntarily participate in this research and give my consent freely. I understand that I can withdraw my participation from the project up to the point of publication, without penalty, and do not have to give any reason for withdrawing.

I understand that all information gathered will be stored securely, and my opinions will be accurately represented. Any data in which I can be clearly identified will be used in the public domain only with my consent.

Participant Signature

Researcher Signature

Date:

Complaints Procedure:

This project follows the guidelines laid out by the Royal College of Art Research Ethics Policy.

If you have any questions, please speak with the researcher. If you have any concerns or a complaint about the manner in which this research is conducted, please contact the RCA Research Ethics Committee by emailing: ethics@rca.ac.uk or by sending a letter addressed to:

The Research Ethics Committee
Royal College of Art
Kensington Gore
London
SW7 2EU

O. SOURCE CODE

<https://github.com/iuliaai/InterSpace.git>

