

# Designing Out Stereotypes in Artificial Intelligence:

Involving users in the personality design of a digital assistant

Dr Jak Spencer<sup>†</sup>

Helen Hamlyn Centre for Design  
Royal College of Art  
London  
jak.spencer@rca.ac.uk

Juliette Poggi

Helen Hamlyn Centre for Design  
Royal College of Art  
London  
juliette.poggi@network.rca.ac.uk

Rama Gheerawo

Helen Hamlyn Centre for Design  
Royal College of Art  
London  
rama.gheerawo@rca.ac.uk

## ABSTRACT

Rapid recent advancements in new technologies and artificial intelligence has led to huge developments in science, technology, medicine and engineering over the last few decades. Recently, developments in the field have tried to create machines that go beyond solving rational, logic based challenges to those that mimic human interaction and in particular emotional intelligence. One of the most mainstream applications of this technology is in digital assistants; tools designed to achieve a result using natural language.

In this paper, we present the findings from a short design research project with a major global accounting firm that aimed to design the personality of a digital assistant. We will present how design methods such as people centred and inclusive design can be used as a tool to build a non stereotypical digital character.

## CCS CONCEPTS

• Human computer interaction • Interaction Design • Systems and tools in interaction design • HCI design and evaluation methods

## KEYWORDS

Digital Assistants, Interaction Design, Human computer interaction,

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [Permissions@acm.org](mailto:Permissions@acm.org)

Goodtechs '18, November 28–30, 2018, Bologna, Italy  
© 2018 Association for Computing Machinery.  
ACM ISBN 978-1-4503-6581-9/18/11 \$15.00  
<https://doi.org/10.1145/3284869.3284897>

## 1 Introduction

Prominent economist Herbert Simon wrote in 1958: “*There are now in the world machines that think, that learn, and that create. Moreover, their ability to do these things is going to increase rapidly until in a visible future the range of problems they can handle will be coextensive with the range to which the human mind has been applied*” [1].

Virtual assistants are machines programmed to ‘think, learn and create’ in a human way. They are designed with a set of characteristics that define their personality.

However, despite recent advancements, creating such intelligent digital personalities remains problematic. These machines tend to portray the stereotypes, judgements and biases of the creators or their culture. As a result, the majority of digital assistants currently on the market show a lack of diversity as the first part of this paper will illustrate.

It is possible and necessary to create personalities out of such stereotype, and design is a key tool. When one of the major accounting firms asked the Helen Hamlyn Centre for Design to design a virtual assistant for their staff, the researchers on this project used people centred design methods to approach the challenge. During a 3-month period that included stakeholder interviews and a co-creation workshop, Helen Hamlyn’s researchers identified the appropriate personality for the future users. The results are presented in this paper and give guide for involving end users in the design of digital personalities

The paper concludes by recommending ways of developing inclusive personalities for digital assistants that put real people at the heart of the design process.

## 2 The problems with giving AI a personality

The seemingly exponential growth of new technology and Artificial Intelligence over the last few decades have led to huge advances in science, technology, medicine and engineering. From an algorithm, capable of recognising a higher percentage of

melanomas than dermatologists [2] to Artificial Intelligence able to create new forms of visual and sound artefacts [3].

In recent years, one area of increasing excitement and activity has been moving from machines with ever increasing ability to solve memorisation or logic challenges to those that show more human traits. Machines are moving towards human-computer interactions that mimic human-human relationships. The field of affective computing opened a new approach to human-computer interaction and led to the desire of designing new types of robots that are emotionally intelligent. Human-Computer empathy is at the centre of this field of research which is now being explored through robots and more mainstream technologies such as Digital Assistants.



Figure 1- Tega Socially Assistive Robot (Westlund, 2017)

One application of this new form of intelligence are Socially Assistive Robots (SAR) like Tega [4] (fig 1). The robot is designed to support interactions with children, with applications in early-literacy education from vocabulary to storytelling. Tega relies on a playful voice and a squashy-soft body to create an empathic connection with its young users.

However, in more mainstream technologies, the most applied variety of AI is that of digital assistants. These are tools designed to “achieve some result by conversing with a machine in a dialogic fashion, using natural language [5]”. From generalists such as Siri, Cortana, Google Assistant and Alexa designed and supported on each specific platform (Apple, Microsoft, Google and Amazon Echo) to more specific functions. From banking to mental health (fig 2), companies are investing in designing virtual personalities that will respond to user’s need, taking virtual assistants to the forefront of interaction design.

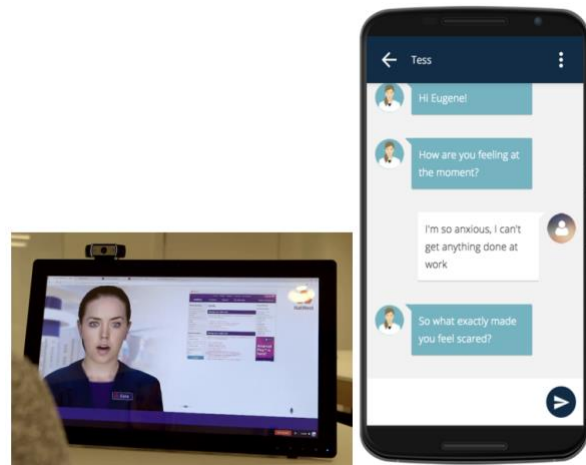


Figure 2 - Banking assistant Cora & mental health Tess

However, despite the excitement and fervent activity in developing emotional robots, there remains many challenges to overcome. One of the biggest of these is the controversial role of playing the ‘creator’. Artificial Intelligence that mimics human relationships have been created by people with a set of preconceived judgments, moralities, ethics and biases. As O’Neil states “algorithms are just opinions embedded in code”[6]. Even as we move towards true intelligent robots that can build themselves, they still use human culture as a source for understanding relationships; stereotypes, discrimination, prejudices and all.

Of course, the difficulty is in giving artificial intelligence an identity. A personality that correlates with your values, without discriminating, yet at the same time an engaging personality that doesn’t bore. Current models of digital assistants are some way from achieving this, with the personality and image of most human-like digital assistants predominantly young, Caucasian women (fig 3).

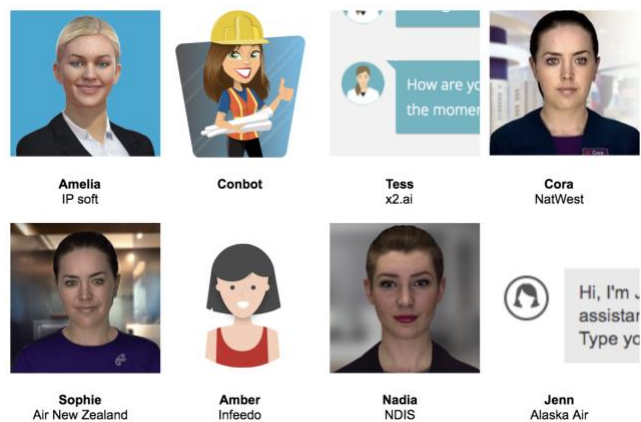


Figure 3 - A selection of current digital assistants

The appearance of such stereotyped digital assistants reveal a design and development process executed in isolation and reflects

a male dominated design industry [7]. Moreover, it enhances the negative perception of the stereotyped role of women in secretarial roles. This is enhanced by the language used by the virtual assistant in general conversation and when responding to certain types of harassment (fig 4).

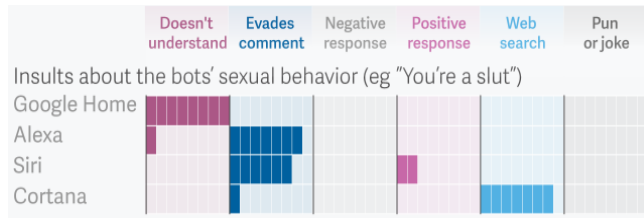


Figure 4 - Bots responses to harassment (Fessler, 2017)[8]

Even popular culture, reflects (or inspires) this stereotypical approach and depiction of technology. Although a few recent movies have had women AI as main characters (Samantha, Theodore’s virtual assistant in *Her*) their personality and role in the plot remain considerably different from the personalities observed in movies with male AI personas (Hal2000, the deadly robot of *2001 Space Odyssey* spaceship).

In fact, despite female personalities being more common as assistants in AI, of the 77 movies from 1927 to 2015 which include AI characters, the vast majority of them are male (fig 5) [9]. The problem persists with the move of ‘always-on’ digital assistants encroaching on every aspect of our lives. “*Someone on TV has only to say, ‘Alexa,’ and she lights up. She’s always ready for action, the perfect woman, never says, ‘Not tonight, dear’*”[10].

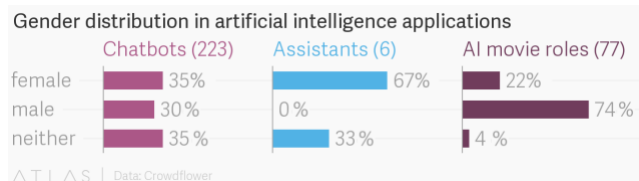


Figure 5 - Gender distribution of AI

Clearly then, developing a digital assistant is more complicated than advancements in the feasibility of the technology and involve challenges which are beyond technical limitations. Models of cultural and social norms have to be challenged and evolved. This is particularly difficult where there are no clear guidelines, and despite the utility of big data in identifying trends and rationalising decisions, a different approach is needed to introduce empathy which will ultimately help building a trustworthy and emotional relationship between the user and the digital assistant.

### 3 Involving users in design

In 2000, the UK government defined Inclusive Design (ID) as products, services and environments that include the needs of the widest number of consumers [11]. It can be defined as a practice, methodology, philosophy or technique, but a key result is that it is internationally recognised and used by governments, industry, designers, policymakers, and social and creative organisations [12]. The genesis of ID is attributed to a paper delivered at the International Ergonomics Association’s 12th Triennial Congress [13] by the founding director of The Helen Hamlyn Centre for Design (HHCD) at the Royal College of Art (RCA) with a British Standard subsequently emerging in 2005 [14].

ID shares a significant part of its ideology with Universal Design and Design for All. All three have their origins in designing for older or disabled people, but different cultural, historical and political factors have affected the way in which these ideals have been interpreted and expressed [15]. Terms such as co-design and co-creation are also in prominent use [16]. Origins can also be traced to Scandinavian Design which aims to enhance quality of life through affordable design solutions that are inclusive rather than exclusive.

At the HHCD, ID has focussed on work with technology companies over the last decade and the effects can be far-reaching, growing social design practice in a variety of sectors, and influencing both economy and environment. Designers are no longer artists in industry but specialist contributors in the knowledge economy. Within business, fostering a culture of innovation is critical to success, as important as mapping out competitive strategies or maintaining good margins [17].

The Helen Hamlyn Centre for Design is the largest and longest running design research centre at the Royal College of Art in London. It specialises in people-centred design to improve life. The challenge, for a small team at the Centre, was to develop the personality of a virtual assistant for a major global firm and move away from the generic, stereotypical design of the current prototype, towards one that would resonate with users. As well as providing more engaging interactions, the aim was to advance more egalitarian ways of designing virtual personalities.

Involving users throughout the development of a new product or service allows a deeper understanding of the psychological and social needs of the user. Design research methods are designed to gain a deep understanding and insight from key stakeholders and user groups. Empathy is crucial for effective design outcomes as it leads to greater levels of inspiration than numbers and measurements alone can provide [18]. “*Empathy can help a designer take a product offering beyond the purely functional, to a new level where the value to the owner is acutely greater, going beyond functionality and usability in product use into the realm of pleasure*”[19].

In a 3-month period from December 2017 to March 2018, the team conducted a series of 1-2-1 interviews with 8 key stakeholders including business analysts, technology leaders and

end users. The aim of these interviews was not only to understand the necessary functionality of a new form or digital assistant, but also to understand the type of interaction desired beyond the traditional user-assistant relationship.

This culminated in a co-creation workshop with 23 future users in which the findings from the interviews were presented. The participants were chosen as diverse members of the organisation from a range of different departments, levels and job functions. Some of them have been working for the companies for over two decades and other just joined the firm a month before the workshop. Variety of ages were present as well, from 25 to 51 years old. Gender wise both sexes were almost equally represented with 10 males and 13 females. All participants were currently living in the UK. Four different teams were randomly formed and were tasked to design the personality for their future virtual assistant using the findings from the research and interviews as outlined below.

#### 4 Designing a personality

Of course, to design a personality, we must first understand the various factors that make one up. The Oxford English Dictionary defines a personality as “the combination of characteristics or qualities that form an individual's distinctive character” [20]. In *The Archetypes and The Collective Unconscious* [21] psychologist, Carl Gustav Jung introduces the concept of the archetypes: recurring themes that can be deduced from mythology and original paintings. More recently, Dr. Carol Pearson, through her research with individuals and organisations created a framework of 12 archetypes that symbolise basic human motivations (fig 6) [22].

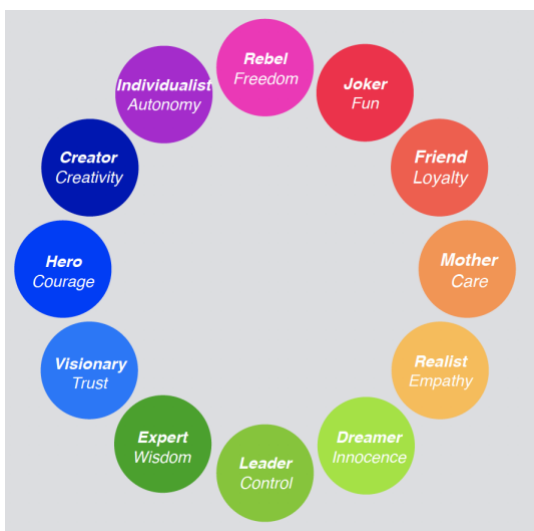


Figure 6 - 12 different personality archetypes

During the workshop, these twelve characters were distilled to a more manageable 4, based on the expectations from the final digital assistant and time available. The idea here was to offer a diverse spread of potential characters that crossed the boundaries of personal, professional, peer and authoritative relationships (fig 7).

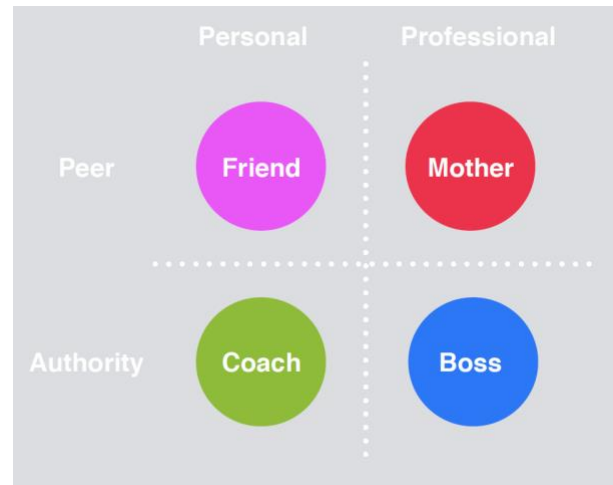


Figure 7 - Characters for development in the workshop

Would the virtual assistant act as your friend and chat about last night's football results? As your mother, who knows you best, offering the most nutritious potential lunch options? Your boss, always giving you something more to work on? Or a coach, to motivate you personally and professionally?

In developing further the characteristics within a personality, 4 key attributes can be identified by research from business and marketing developed to best create and market brands [23].

*Purpose & Story:* Why does it exist in the first place?

*Person:* Who does it sound like?

*Tone of Voice:* How do you describe it?

*Language:* What are the words and phrases used?

The purpose is of course a necessary attribute in creating a tool with relevance to users. There is a proliferation of digital assistants currently on the market, and therefore, if users are to engage with the tool, it must serve a real need. The stakeholder interviews helped to identify specific issues that were not currently being served by existing tools. If the purpose is the future mission of the digital assistant, there is also the need to understand what is the character's past. It is important for the character created to understand its own self, creating its story to make it relate to a set of characteristics.

Once the purpose and story are clear, defining who it should sound like come next. Participants were asked to map out the core values of their character and the motto they would live their life

by. Are they creative and spontaneous, loyal, inspirational or direct? Would you call them up for a good time, or when you are in trouble or need advice? Would you go on holiday with them, or to a conference?

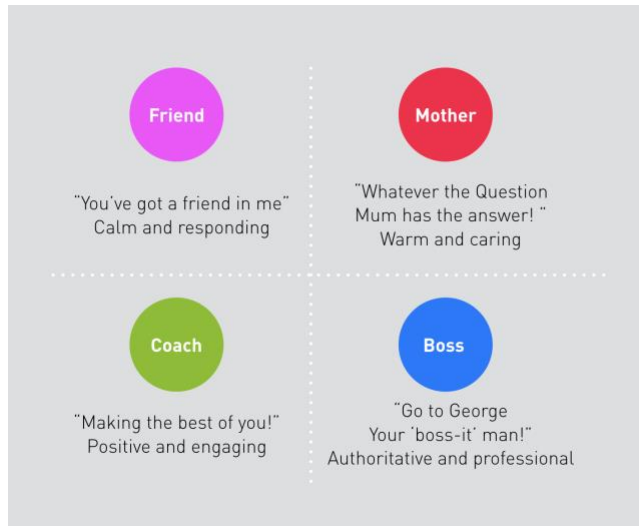


Figure 8 – The final characters’ claim and tone of voice

Once the person archetype was developed with the values that the personality embodied, the tone of voice was defined to understand how the personality will communicate its values. This is important in developing the relationship with the user. Will the personality be funny or serious? Formal or causal? Respectful or irreverent? Enthusiastic or matter of fact?

Finally, the language portrays how the tone of voice is expressed. It involves phrase construction, words and pronouns. It is particularly important in the hierarchical balance of the relationship between the user and the digital assistant. Is the assistant there to serve, or is the relationship more equal? The language will portray this, for example: “Do *you* need help” suggests a proactive engagement with the user, versus “May *I* help you” puts the assistant’s agency above the user. The different teams role-played an example of the conversation using the language of their character.

Each team presented each stage of the character creation to the rest of the audience. At the end of the workshop every character was evaluated by the future users according to three parameters:

- 1- Which character they trust the most?
- 2- Which character they feel more comfortable with?
- 3- Which character is more relevant to the company?

## 5 The creation of a new digital assistant

In framing the creative workshop like this, the users created 4 distinct personalities for their future digital assistant. These personalities moved beyond the generic, stereotyped versions of

previous iterations, towards diverse characters with unique attributes.

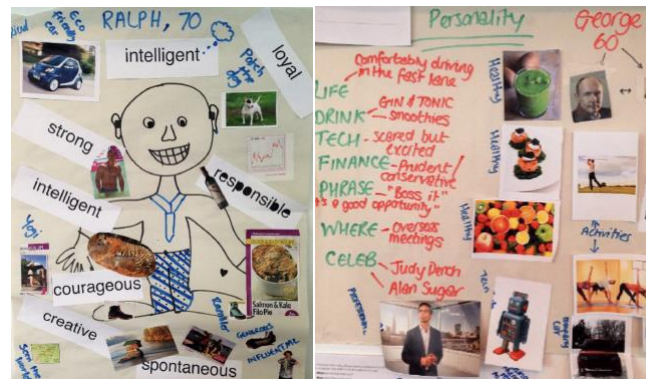


Figure 9 - Example mood boards for new characters

The 4 characters were quite defined from each other and had some common characteristics such as being all British and all allocating social time in their day. However, the most surprising and relevant result was that, out of the 4 personalities created, 3 were over the age of 50, with an even gender split.

The ‘friend’ personality was deemed by all participants to be the most suitable for the application in a work environment and offer egalitarian peer support. However, unlike a friend of similar age or background, users wanted an older, wiser friend to offer them support. Characters such as Judy Dench or David Attenborough were popular to mimic in this role, as they embodied the kind, sincere and knowledgeable traits that people looked for in a supporting assistant. Someone they could confide in and offer humour and support when needed.

The initial research has suggested a need for more diverse characters in digital assistants, and using the inclusive design co-creation methodology has allowed a number of different personalities to be developed. Involving real users from diverse backgrounds, be it from the workplace, or cultural or economic backgrounds leads to new ideas that would otherwise not be considered by select groups of developers. This research led to the design of a set of ‘personality guidelines’ that the company can give to their coders and developers to incorporate the new personality in their digital assistant personality. This is currently in the prototyping phase.

Whilst, of course, design in this way does not preclude entirely negative stereotypes and prejudices to filter through –by involving real users in the design of digital personalities, an emotional link is made between the user and machine during such a creative process. Stereotypes can be addressed at the outset and clearer definitions around the role the digital assistant will take on can be explored. Moving from subservient assistant to peer relationships such as ‘friend’, inspirational such as ‘coach’ or authoritative such as ‘boss’. It also ensures that interaction decisions are not made

by coding developers, detached from real users. It also starts to humanise a process often seen as too technology led to create personalities that are playful, humorous, relevant and engaging. Users feel involved in the creation of new technologies and thus, are more engaged when the product launches.

Of course, we must also recognise some of the limits of the work. This was a short piece of work with a limited number of users for a global accounting firm. Stereotyped digital personalities is a huge subject area that requires considerable further work, and this research only scratches the surface. However, it shows that, even over a short period of time, with just a limited number of real users, 4 very diverse and distinct personalities were created. In this way, it shows the huge potential that combining more people-centred design approaches with new technologies and big data can lead to innovative and impactful results. This is something we must surely see more of in the coming years as new technologies continue to advance at ever-increasing speed – we must not forget about the user.

## ACKNOWLEDGMENTS

The authors would like to thank Tata Consultancy Services for their continued support as well as the participants of the workshop for their engagement and time.

## REFERENCES

- <sup>1</sup> H. A. Simon and A. Newell (1958). 'Heuristic Problem Solving: The Next Advance in Operations Research', *Operations Research*, Vol. 6 (Jan-Feb 1958), 1-10.
- <sup>2</sup> A. Esteva et al. (2017). "Dermatologist-level classification of skin cancer with deep neural networks", *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.
- <sup>3</sup> Sony CSL (2018). Flow Machines example: <http://www.flow-machines.com/>
- <sup>4</sup> J.M.K Westlund. (2017). Making new (robot) friends: Understanding children's relationships with social robots. <https://medium.com/mit-media-lab/making-new-robot-friends-b68ae16809a3>
- <sup>5</sup> Dale, R. (2016). The return of the chatbots. *Natural Language Engineering*, 22 (5), 811–817
- <sup>6</sup> O'Neil. (2016). *Weapons of Math Destruction*. New York : Crown.
- <sup>7</sup> S. Reimer (2016) 'It's just a very male industry': gender and work in UK design agencies. *Gender, Place & Culture*, 23:7, 1033- 1046
- <sup>8</sup> L. Fessler. (2017). We tested bots like Siri and Alexa to see who would stand up to sexual harassment. *Quartz Magazine*
- <sup>9</sup> T. Schnoebelen. (2016). The gender of artificial intelligence. *Future Eight*. <https://www.figure-eight.com/the-gender-of-ai/>
- <sup>10</sup> S. Sage. (2017). Quoted in 'Alexa, Where Have You Been All My Life?'. P. Green. *The New York Times*.
- <sup>11</sup> Department of Trade and Industry. (2000) *Foresight: Making the Future Work for You*. UK: Department of Trade and Industry.
- <sup>12</sup> Clarkson J, Coleman R, Keates S, Lebbon C. (2003) (eds.) *Inclusive Design – Design for the Whole Population*. London: Springer-Verlag.
- <sup>13</sup> R. Coleman. (1994) The Case for Inclusive Design – An Overview, *Proceedings of the 12th Triennial Congress, International Ergonomics Association and the Human Factors Association of Canada*.
- <sup>14</sup> British Standards Institute (2005). *BS 7000-6; Design management systems. Managing inclusive design*. Guide. UK.
- <sup>15</sup> T. Vavik and R Gheerawo. (2009) The Challenges in Universal Design. In Vavik T (ed.) *Inclusive Buildings, Products & Services: Challenges in Universal Design*. Trondheim: Tapir Academic Press.
- <sup>16</sup> Sanders E B N, Stappers P J. (2014) Editorial. In *CoDesign: International Journal of CoCreation in Design and the Arts*, 10:1.
- <sup>17</sup> Kelley T. (2006). *The Ten Faces of Innovation – Strategies for Heightening Creativity*. London: Profile Books.

---

<sup>18</sup> C. McGinley & H. Dong (2015). Designing with Information and Empathy: Delivering Human Information to Designers, *The Design Journal*, 14:2, 187-206, DOI: 10.2752/175630611X12984592780005

<sup>19</sup> P. Jordan. (1997). 'Human factors for pleasure in product use'. *Applied Ergonomics*, 29(1), 25–33.

<sup>20</sup> Oxford English Dictionary. (2015). Oxford University Press

<sup>21</sup> Jung, C. G. (1981). *The archetypes and the collective unconscious* (R. F. C. Hull, Trans.). Princeton, NJ: Princeton University Press. (Originally published 1954)

<sup>22</sup> M. Mark and C. Pearson. (2001). *The hero and the outlaw: Building extraordinary brands through the power of archetypes*. New York: McGraw-Hill.

<sup>23</sup> K. Lee. (2014). The Best Examples, Questions, and Guides to Find Your Social Media Marketing Voice. *Fast Company Magazine*