



Pelikan, Michal Lech (2026) *Making digital avatars appear more alive: reverse correlation of humanlike behaviour and the uncanny valley*. PhD thesis.

<https://theses.gla.ac.uk/86188/>

Copyright and moral rights for this work are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This work cannot be reproduced or quoted extensively from without first obtaining permission from the author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given

Enlighten: Theses

<https://theses.gla.ac.uk>

research-enlighten@glasgow.ac.uk



University
of Glasgow

Making Digital Avatars Appear More Alive: Reverse Correlation of Humanlike Behaviour and the Uncanny Valley

Michal Lech Pelikan

Submitted in fulfilment of the requirements
for the Degree of Doctor of Philosophy

School of Computing Science
School of Psychology and Neuroscience
College of Science & Engineering
University of Glasgow

March 2026

Acknowledgements

Firstly, I would like to thank the SocialAI CDT directors for giving me the opportunity to be part of the programme, and all members of the CDT team who have supported both me and my work during my time at the University of Glasgow. I am especially grateful to Monika Harvey for her continued support, particularly in the final stages of the project. Her guidance and encouragement during this period made a real difference and helped me navigate the last stretch of the thesis.

I extend my sincere thanks to Professor Philippe Schyns. I am truly grateful for your mentorship and guidance throughout this project, as well as the trust and support you gave me in exploring my own ideas through this research.

My thanks also go to all members of the Schyns Lab for their support, particularly through valuable feedback on presentations and ongoing discussions that helped refine this work. I am especially grateful to Oliver Garrod and Chen Zhou for the insightful conversations and technical advice, which were instrumental in helping me overcome key challenges throughout the project.

I am also grateful to Professor Faisal Mushtaq for his support over the years. The opportunities and guidance he provided allowed me to develop key skills and gain valuable experience, both of which played an important role in my progression onto and completion of the PhD programme.

To my friends and family, thank you for your constant support over the years. I know that, at times, I struggled to maintain a healthy work–life balance, and that meant sacrificing valuable time with you. I am truly grateful for your understanding and for always being there for me throughout this journey. As I close this chapter, I look forward to making up for all the time we have missed together.

Drodzy Rodzice, nie macie pojęcia, jak bardzo jestem Wam wdzięczny za to, że jesteście i zawsze byliście przy mnie. Być może nie zdajecie sobie w pełni sprawy, jak ogromny wpływ mieliście na to, gdzie jestem dzisiaj i jak wiele Wam zawdzięczam.

Mimo że nie zawsze było nam łatwo, zawsze robiliście wszystko, co w Waszej mocy, aby niczego nam nie brakowało. Często wiązało się to z wieloma wyrzeczeniami, długimi rozłąkami i latami ciężkiej pracy za granicą, aby zapewnić nam jak najlepszą przyszłość.

Tato, dziękuję Ci za wszystkie lata poświęcenia i ciężkiej pracy, które włożyłeś w to, abyśmy mogli mieć lepsze życie. Mamo, dziękuję Ci szczególnie za ogromną odwagę, jaką wykazałaś, opuszczając naszą ojczyznę i przeprowadzając się do Anglii z nadzieją na lepszą przyszłość dla nas.

Mając jedenaście lat, nie znając języka i zaczynając wszystko od zera, potrzebowałem czasu, aby w pełni zrozumieć i docenić tę decyzję. Dziś nie zmieniałbym jej za nic. Gdyby nie ten jeden trudny krok, nie byłbym teraz w tym miejscu i nie osiągnąłbym tego wszystkiego bez Waszego wsparcia.

Kocham Was i z całego serca dziękuję za wszystko.

Lastly, from the bottom of my heart, I would like to thank the love of my life, Dr Layla Kouara. I could not have done this without you. You have been by my side at every step of this journey, supporting me and helping me grow into the best version of myself. Through all the stressful days and sleepless nights, both the good and the difficult moments, you were always there for me, and I am so grateful to have you in my life.

None of this would have been possible without you by my side, and as this chapter comes to an end, I look forward to everything that lies ahead for us.

I love you dearly, and thank you so much for everything.

Declaration

I declare that, except where explicit reference is made to the contribution of others, this dissertation is the result of my own work and has not been submitted for any other degree at the University of Glasgow or any other institution. I hereby give my permission for this thesis to be shown to other University of Glasgow students and to be distributed in an electronic format.

The present thesis was supported by the UKRI Centre for Doctoral Training in Socially Intelligent Artificial Agents, Grant Number EP/S02266X/1.

Contents

1	General Introduction	17
1.1	History of digital avatars in media	17
1.2	Technological progress in rendering and animation	20
1.3	The uncanny valley	27
1.4	Motion and realism	31
1.5	Modelling human perception	32
1.6	The present thesis	36
2	Unreal Engine Animation Framework	39
2.1	Chapter Overview	39
2.2	Motivation and Requirements	39
2.3	Architecture Overview	42
2.4	CSV Schema	43
2.5	Value ranges and data conventions	44
2.6	Temporal sampling and frame rate	44
2.7	Handling multiple clips in a single file	44
2.8	Recommended practices for data management	45
2.9	Importing Data into Unreal Engine	46
2.10	Animation Capture Settings	48
2.11	Rendering and Post-Processing	49

2.12	Further Automation	49
2.13	Technical Considerations and Limitations	49
2.14	Chapter Summary	50
3	Validating MetaHuman Avatars: A Mixed Methods Benchmark (Study 1a & 1b)	52
3.0.1	Chapter Overview	52
3.1	Study 1a: Quantitative Benchmark of MetaHuman and GFG Avatars	53
3.1.1	Introduction	53
3.1.2	Methods	61
3.1.3	Results	70
3.1.4	Discussion	74
3.1.5	Conclusion	80
3.2	Study 1b: Qualitative Evaluation of MetaHuman and GFG Avatars	81
3.2.1	Study Overview	81
3.2.2	Introduction	82
3.2.3	Methods	88
3.2.4	Results	94
3.2.5	Discussion	106
3.2.6	Conclusion	110
4	Reverse Correlation of Natural Idle Facial Motion (Study 2)	112
4.0.1	Chapter Overview	112
4.1	Introduction	112
4.1.1	Uncanny Valley and the Role of Motion in Human-likeness	114
4.1.2	Reverse Correlation in Social Perception Research	117
4.1.3	Reverse Correlation, Idle Motion, and the Present Study	119
4.2	Methods	120

4.2.1	Participants	120
4.2.2	Materials	120
4.2.3	Stimuli	121
4.2.4	Procedure	128
4.3	Results	130
4.3.1	Data Analysis	130
4.3.2	Action Units	130
4.3.3	Head and Eye Motion	133
4.3.4	Validation	133
4.4	Discussion	137
4.4.1	MetaHumans and Reverse Correlation	139
4.4.2	Perceived Naturalness of Idle Facial Motion	141
4.4.3	Implications for Uncanny Valley Research	143
4.4.4	Applied and Practical Implications	144
4.4.5	Future Directions	145
4.5	Conclusion	147
5	Motion Naturalness and the Uncanny Valley in Digital Avatars (Study 3)	148
5.0.1	Chapter Overview	148
5.1	Introduction	149
5.1.1	The Uncanny Valley: From Appearance to Perceptual Mismatch	149
5.1.2	Motion as a Critical but Under-Isolated Component of the Uncanny Valley	150
5.1.3	Establishing a High-Fidelity Baseline: MetaHumans as Controlled Stimuli (Chapter 3)	151

5.1.4	Modelling Idle Motion Naturalness Through Reverse Correlation (Chapter 4)	152
5.1.5	Testing Motion Naturalness Within the Uncanny Valley Framework	153
5.1.6	Hypotheses and Broader Theoretical Contribution	153
5.2	Methods	155
5.2.1	Participants	155
5.2.2	Materials	155
5.2.3	Procedure	158
5.2.4	Data Analysis	159
5.3	Results	160
5.3.1	Human-likeness	160
5.3.2	Eeriness	161
5.3.3	Familiarity	161
5.3.4	Realism	162
5.4	Discussion	163
5.4.1	Eeriness and Dynamic Cue Mismatch	164
5.4.2	Motion Quality Beyond Eeriness	165
5.4.3	Methodological Contribution	166
5.5	Conclusion	167
6	General Discussion	168
6.1	Contributions of Each Chapter	168
6.2	What the Thesis Changes About Motion and the Uncanny Valley	171
6.3	Extending Reverse Correlation to Idle Facial Motion	172
6.4	Methodological Contributions	173

6.5 Practical Implications for Digital Human Creation 174

6.6 Limitations 175

6.7 Future Directions 177

6.8 Conclusion 180

List of Figures

1.1 Examples of digital human characters across five decades, comparing real-time (RT) and pre-rendered (PR) rendering pipelines. Real-time characters are generated instantly by a graphics engine and are designed for interactive media such as video games, whereas pre-rendered characters are computed offline without real-time constraints, typically for film production. The figure illustrates the progression of digital human realism from early polygonal models to contemporary high-fidelity characters with physically based rendering and advanced facial animation, including (A) Mike the Talking Head (1988; RT), (B) Rendez-vous in Montreal (1987; PR), (C) Tomb Raider (1996; RT), (D) Toy Story (1995; PR), (E) Uncharted: Drake’s Fortune (2007; RT), (F) Final Fantasy: The Spirits Within (2001; PR), (G) Uncharted: A Thief’s End (2016; RT), (H) Blade Runner 2049 (2017; PR), (I) The Last of Us: Part I (2022; RT), and (J) Blue Dot (2024; PR). 19

1.2 Effect of subsurface scattering (SSS) on digital skin rendering. The left column shows rendering with subsurface scattering enabled, and the right column shows rendering with subsurface scattering disabled. (A) shows the full face, (B) shows a closer view of facial skin, and (C) shows a close-up of an ear. When SSS is enabled, light penetrates the skin, scatters beneath the surface, and exits at nearby points, producing softer shading, subtle colour variation, and a more natural sense of depth. This effect is particularly visible in thinner regions such as the ear, where transmitted light creates a reddish glow (Panel C). When SSS is disabled, skin appears more opaque, flatter, and more plasticky due to the absence of subsurface light transport. 22

1.3	Overview of selected Facial Action Units (AUs) defined within the Facial Action Coding System (FACS). FACS decomposes facial behaviour into anatomically based components corresponding to specific muscle contractions or relaxations. The upper panel shows action units primarily involving the upper face, such as brow raising and eyelid movements, while the lower panel shows action units involving the lower face, including lip, jaw, and chin movements. Each action unit represents a distinct, observable facial movement. Reproduced from De la Torre et al. (2015) with permission © [2015] IEEE.	24
1.4	Example of a MetaHuman rendered in Unreal Engine using physically based materials and real-time lighting. The image illustrates high-resolution facial geometry, detailed skin texture with subsurface scattering, realistic eye shading, and strand-based hair simulation. Platforms such as MetaHuman enable industry-level visual fidelity within an openly accessible real-time engine, reducing the technical and financial barriers to creating photorealistic digital humans for research and production contexts.	27
1.5	Original illustration of the uncanny valley hypothesis proposed by Mori (1970). The vertical axis represents affinity, and the horizontal axis represents increasing human likeness. As artificial agents become more humanlike, affinity increases until a critical point at which near-human entities elicit a sharp decline in emotional response, forming the so-called uncanny valley. The dashed line indicates that movement can intensify this effect, producing a deeper valley compared with still entities. Examples such as prosthetic hands, corpses, and zombies are positioned within the valley, whereas healthy humans occupy the highest level of affinity. Adapted with permission from Mori et al. (2012); © 2012 IEEE. . .	28
1.6	Illustration of the reverse correlation method applied to facial perception. In the stimulus generation stage (i), random white noise is added to a neutral base face to create slightly varied stimulus images. Participants categorise each noisy face according to emotional judgments (e.g., happy, anger, sad). In the analysis stage (ii), noise patterns associated with each response category are averaged across trials to estimate the visual features influencing those judgments. The resulting classification images represent observers' mental representations of different emotions (iii), shown here for Western and East Asian participants. Adapted from Jack & Schyns (2017) with permission.	33

1.7	Example of a facial stimulus generated using the Generative Face Grammar (GFG; Yu et al., 2012). Although highly controllable and methodologically influential in reverse correlation research, the visual fidelity reflects the graphics standards available at the time of its development.	36
2.1	Example Animation Blueprint setup for driving MetaHuman facial controls from imported CSV data. The AnimGraph shows Control Rig nodes with exposed input pins corresponding to Action Unit (AU), eye, and head rotation variables. These variables are updated frame-by-frame from the DataTable and routed into the MetaHuman Control Rig, enabling real-time application of externally generated motion parameters to the character’s facial mesh.	47
2.2	Frame-based iteration of animation data within the Animation Blueprint Event Graph. (A) A Set Timer by Event node is configured to fire at 60 frames per second, triggering the FireAnimFrame event at a fixed rate. (B) Each time the event fires, the system reads the corresponding row from the animation DataTable, retrieving the Action Unit (AU) activation values for the current frame. (C) These values are assigned to the Animation Blueprint variables and passed to the MetaHuman Control Rig, updating AU activations and facial pose in real time. Collectively, this setup enables the DataTable to be read sequentially, one row per frame, ensuring frame-accurate playback of externally generated motion data at the specified framerate.	48
3.1	Stimulus characters used in the experiment. Six identities were first generated using the Generative Face Grammar (GFG) framework and then transferred to the MetaHuman system using the Mesh to MetaHuman tool in Unreal Engine. (A) shows the original GFG avatars, while (B) shows the corresponding MetaHuman counterparts reconstructed from the GFG meshes. Each row represents a matched identity pair, with three male and three female characters spanning three broad ethnic groups (Caucasian, East Asian, and African). The MetaHuman characters were subsequently refined in MetaHuman Creator by manually selecting the closest matching hair style and adjusting facial and eye textures to approximate the appearance of the original GFG identities as closely as possible. The original GFG characters comprised the head and neck only, whereas the corresponding MetaHuman characters included part of the upper torso as MetaHumans come with a full body rig. The characters were framed so that the heads were equally visible across both avatar frameworks.	64

3.2	Comparison of mesh geometry between the original Generative Face Grammar (GFG) character and its reconstructed MetaHuman (MH) counterpart for the Caucasian male identity. Columns show the GFG mesh, the reconstructed MetaHuman mesh, and an overlay of both meshes for visual comparison. Rows A and B show wireframe views highlighting mesh topology from frontal and lateral perspectives, while rows C and D show the corresponding clay renders illustrating overall surface geometry. The comparison demonstrates that the global facial structure of the identity is largely preserved during the Mesh to MetaHuman transfer process.	66
3.3	Localised topology differences between the GFG and MetaHuman meshes. The top row (A-C) shows the original Caucasian male GFG mesh, while the bottom row (D-F) shows the corresponding Caucasian male MetaHuman reconstruction generated using the Mesh to MetaHuman pipeline. Panels (A) and (D) show the nasal region, (B) and (E) show the close-up of the frontal facial mesh, and (C) and (F) show a close-up of the eye region. Compared to the GFG mesh, the MetaHuman mesh introduces additional anatomical detail, including more defined nostril geometry, layered eyelid structure, and tear duct regions around the eye. These differences arise from the higher resolution topology used by the MetaHuman framework and illustrate how local geometric detail is added during the identity transfer process while preserving the general overall facial structure.	67
3.4	Visual comparison of surface detail between the original GFG avatar and its MetaHuman counterpart for the Caucasian male identity. (A) shows the GFG character, while (B) shows the reconstructed MetaHuman version generated using the Mesh to MetaHuman pipeline. Although the overall facial identity is preserved, several differences arise from the distinct material and hair systems used by the two frameworks. In addition to the previously described mesh topology differences, the GFG model represents eyebrows as a two-dimensional texture mapped onto the facial surface and does not include eyelashes. Whereas the MetaHuman character uses strand-based systems with individually modelled three-dimensional hairs for both eyebrows and eyelashes, along with higher resolution skin and eye materials. These differences contribute to the visual variations between the two avatar representations despite efforts to closely match the original identity.	69

3.5	Boxplots showing participant ratings of avatars generated with the Generative Face Grammar (GFG) framework and Unreal Engine MetaHumans (UE) across four perceptual dimensions: human-likeness, eeriness, familiarity, and realism. The horizontal line inside each box shows the median rating, the diamond shows the mean rating, the box shows where the middle 50% of ratings fall, and the whiskers show the spread of most of the remaining ratings. Individual participant ratings are shown as transparent points. Ratings were made on a 7-point Likert scale. UE avatars were rated as significantly more human-like, familiar, and realistic than GFG avatars, whereas GFG avatars were rated as significantly less eerie. Asterisks indicate significance levels from paired samples t-tests following Holm-Bonferroni correction for multiple comparisons(*p < .05, ***p < .001).	72
3.6	Perceived identity similarity between paired GFG and MetaHuman avatars for each character identity. Bars show the mean similarity rating on a 7-point scale (1 = “Not at all the same person”, 7 = “Definitely the same person”), with error bars representing 95% confidence intervals. The dashed vertical line indicates the neutral midpoint of the scale (4). Asterisks denote significant differences from the midpoint based on one-sample t-tests (**p < .01, ***p < .001). . . .	74
3.7	Stimulus characters used in the experiment. Six identities were first generated using the Generative Face Grammar (GFG) framework and then transferred to the MetaHuman system using the Mesh to MetaHuman tool in Unreal Engine. (A) shows the original GFG avatars, while (B) shows the corresponding MetaHuman counterparts reconstructed from the GFG meshes. Each row represents a matched identity pair, with three male and three female characters spanning three broad ethnic groups (Caucasian, East Asian, and African). The MetaHuman characters were subsequently refined in MetaHuman Creator by manually selecting the closest matching hair style and adjusting facial and eye textures to approximate the appearance of the original GFG identities as closely as possible.	91

3.8	Mean ratings of socio-cognitive attributes for GFG and MetaHuman (UE) avatars across four dimensions: emotional capacity, distinct personality, person-likeness, and moral concern. Ratings were made on a 5-point scale, with higher values indicating stronger agreement that the avatar possessed the given attribute. Error bars represent ± 1 standard error of the mean. Asterisks indicate significant differences between frameworks (* $p < .05$, *** $p < .001$). MetaHuman avatars were rated significantly higher than GFG avatars on emotional capacity, person-likeness, and moral concern, while no significant difference was observed for distinct personality.	95
3.9	Distribution of pronoun attributions (%) for GFG and MetaHuman (UE) avatars across character identities. Stacked bars represent the proportion of responses assigned to each pronoun category (he, she, it, they, other) for each avatar. Across both frameworks, pronoun use generally followed perceived gender cues; however, MetaHuman avatars elicited more consistent use of gendered pronouns, whereas GFG avatars showed greater variability and more frequent use of non-gendered or objectifying terms (e.g., “it” and “they”).	97
4.1	Experimental setup used for stimulus presentation and response collection. Participants were seated at a fixed viewing distance using a chin rest to maintain a consistent visual angle (approximately 8°) while viewing stimuli on a 24” monitor. Responses were provided via a keyboard, with the laptop positioned out of sight to minimise distraction and ensure standardised viewing conditions across participants.	121
4.2	MetaHuman identities used to generate the stimulus set. The full set comprised 15 characters (8 male, 7 female) varying in age and ethnicity to capture a broad range of facial identities. All characters were created using Unreal Engine’s MetaHuman framework and served as the base identities for the animated stimuli. This diversity was included to ensure that the results were not driven by a specific face or demographic category.	122

4.3	Parameterisation of facial action unit (AU) activation curves used for animation synthesis. The temporal profile of each AU was defined by seven parameters: peak amplitude, peak latency, onset latency, offset latency, acceleration, deceleration, and sustain. These parameters were generated from a set of random variables (RV1-RV6) and combined using a function adapted from Yu et al. (2012), with the addition of a sustain component. The sustain parameter controls the duration for which the AU remains at its peak activation before returning to the baseline. This formulation defines the timing and shape of each AU activation curve, which was then used to drive facial motion in the generated animations. .	123
4.4	Illustration of action unit (AU) activation curve parameterisation and example curves. (A) Schematic of a single AU activation curve showing how the control point parameters define its temporal structure, including onset latency, peak latency, offset latency, peak amplitude, and the acceleration and deceleration phases. (B) Example activation curves demonstrating how variation in parameter values may produce different temporal profiles of AU activation.	124
4.5	Temporal structure of the procedurally generated animation components in the final stimuli. Each stimulus consisted of multiple layers of motion generated over a 4-second duration. Head motion, eye saccades, and blinks were present continuously throughout the full 4-second animation, while facial action unit (AU) activity was embedded as a shorter 2-second segment that was presented twice within the sequence.	127
4.6	Temporal structure of validation stimuli. Each validation animation lasted 6 seconds, with continuous head motion, eye saccades, and blinks present throughout the full duration. Facial action unit (AU) activity was embedded as a single 2-second segment occurring midway through the animation (2-4 seconds), with no repetition.	128
4.7	Frequency of facial action units (AUs) significantly associated with perceived naturalness across participants. Bar plots show the number of participants (out of 18) for whom each AU was significantly positively correlated with naturalness ratings ($p < .05$), grouped by rating category: low (1-2), mid (3-5), and high (6-7). AU labels correspond to the naming conventions of the MetaHuman facial rig controls used to drive animation. AUs were treated as bilateral, such that if either the left or right lateralised variant of an AU was significant for a given participant, the AU was counted as significant overall. Dashed lines indicate the threshold of 4 participants (22%), which was used to determine AU inclusion in subsequent animation synthesis models.	132

4.8	Mean congruence accuracy of reverse-correlated idle motion models across naturalness categories. Bars show the mean classification accuracy (%) with which naïve observers (n = 10) correctly identified the intended naturalness category (Low, Mid, High) of the generated animations, along with the overall accuracy across all conditions. Congruence accuracy was defined as the proportion of trials in which an observer’s rating category matched the intended category. Error bars represent the standard error of the mean (SEM) across participants. The dashed line indicates chance performance (33.3%) corresponding to three response categories.	134
4.9	Confusion matrix showing the percentage of participant judgements for each naturalness category. Diagonal values indicate correct classifications, whereas off-diagonal values indicate misclassification across categories.	135
4.10	Representative frames from the three most consistently classified validation animations in each intended naturalness condition (High, Mid, Low), shown at approximately peak facial action unit activation. Animations are listed in order of appearance. Percentages indicate the proportion of naïve observers whose ratings matched the intended naturalness category of each animation. High naturalness: 84_Wallace (90%), 42_Kioko (90%), 23_Irene (80%). Mid naturalness: 52_Layla (70%), 34_Kendra (70%), 22_Irene (70%). Low naturalness: 02_Bernice (80%), 07_Cooper (90%), 19_Irene (80%).	137
5.1	MetaHuman characters used as base identities for the motion stimuli. The stimulus set comprised six MetaHuman characters arranged by gender (rows) and ethnicity (columns), including Caucasian, East Asian, and African identities, with one male and one female character representing each category.	156
5.2	Mean ratings (1-7 scale) of human-likeness, eeriness, familiarity, and realism as a function of idle motion condition (Still, Low Naturalness, High Naturalness). Error bars represent 95% within-subject confidence intervals (Cousineau-Morey corrected). Asterisks indicate significant pairwise comparisons (Bonferroni corrected): *p < .05, **p < .01, ***p < .001.	163

Chapter 1

General Introduction

1.1 History of digital avatars in media

In the early years of computer graphics, digital faces were built from basic shapes and flat textures. They were remarkable for their time, but clearly artificial. During the 1980s and 1990s, improvements in computing power and software allowed artists to model more realistic facial structure and lighting (Foley, 1996). Even so, these early computer-generated faces were limited by the available hardware and rendering methods. Although the overall shape of a face could look convincing, the surface often appeared smooth or plastic, and the movement lacked the small variations that make real people seem alive. As a result, early digital humans rarely felt like believable individuals.

A major shift occurred when films began combining computer-generated characters with live action footage. Jurassic Park, for instance, showed that digital creatures could be convincingly integrated into live action scenes, with realistic lighting and motion (Prince, 2011). It demonstrated that digital creatures could feel grounded in the real world, raising audience expectations for realism and laying the groundwork for later attempts at believable digital humans. However, achieving the same effect with digital faces proved far more difficult, because audiences have deep perceptual expertise in reading real human faces and are highly sensitive to even the smallest inconsistencies in skin texture, facial structure, timing, and expression (Kilts et al., 2003; Maurer et al., 2002; Troje & Bülthoff, 1996). Unlike dinosaurs or fantastical creatures, human faces are stimuli that observers encounter every day, and they are processed with exceptional perceptual precision (Haxby et al., 2000).

When a digital character is shown in close-up, especially as a human lead, the viewer becomes far more sensitive to small errors. *Figure 1.1* illustrates examples of digital human characters across decades, distinguishing between real-time characters used in interactive media

such as games and pre-rendered characters produced offline for film. A well-known example is *Final Fantasy: The Spirits Within* (2001; Fig.1.1F), which attempted to portray photorealistic human characters by using advanced skin texture and early, crude motion capture (only used for body motion) but received criticism because the faces appeared stiff or emotionally flat (Pollick, 2010). These early examples demonstrated how realism is not only about how photorealistic a character looks, but also about how the character behaves, and if this behaviour matches its appearance. This means that in order to achieve realism, the challenge is not just to copy the anatomical structure of a face, but to recreate the combined effect of the internal structure and motion that makes a person appear alive (Kokkinara & McDonnell, 2015).

This sensitivity to small facial cues matters because faces are one of the main sources of social information. In everyday life, people form impressions from faces quickly, sometimes after brief exposure, and these impressions can include whether someone seems trustworthy, threatening, confident, or physically strong (Oosterhof & Todorov, 2008; Willis & Todorov, 2006). These rapid judgments shape real social outcomes, including who we choose to approach, avoid, follow, or punish, even when we believe we are being objective (Oosterhof & Todorov, 2008). While this does not mean the judgments are accurate, it means that the human visual system is tuned to extract social meaning from facial cues, and it often does so automatically.

In a study using 3D face models, Oosterhof and Todorov (2008) asked participants to rate emotionally neutral faces on multiple trait dimensions. Based on these ratings, they showed that many different trait judgments from neutral faces can be organised using two broad dimensions. They called these dimensions valence and dominance. In simple terms, valence captures whether a face gives an impression of good or bad intent, which often resembles cues linked to approach and avoidance. Dominance captures whether a face gives an impression of physical strength or weakness, which relates to perceived ability to cause harm. According to this model, specific impressions such as trustworthiness and threat can be understood as combinations of these two underlying dimensions. They also showed that these dimensions are linked to subtle aspects of face shape and configuration, rather than obvious expressions. Using a data-driven 3D face model, they manipulated facial shape to demonstrate how small structural changes systematically altered social judgments. This suggests that even small changes in the facial structure of digital humans can shift how a character is judged, even when the face is emotionally neutral.

A single face can communicate many kinds of information at once, including identity, age, gender, health, emotional state, and likely intention (Jack et al., 2016; Lick et al., 2013; Russell et al., 2016; Todorov et al., 2015; Tsantani et al., 2021; Zhan et al., 2022). Because so much meaning is carried in one stimulus, faces provide a powerful tool for studying social cognition. Social cognition refers to the mental processes that allow people to perceive, interpret, and respond to social information. It includes forming impressions of others, predicting how they might behave, and adjusting one's own behaviour accordingly. Faces are therefore central to

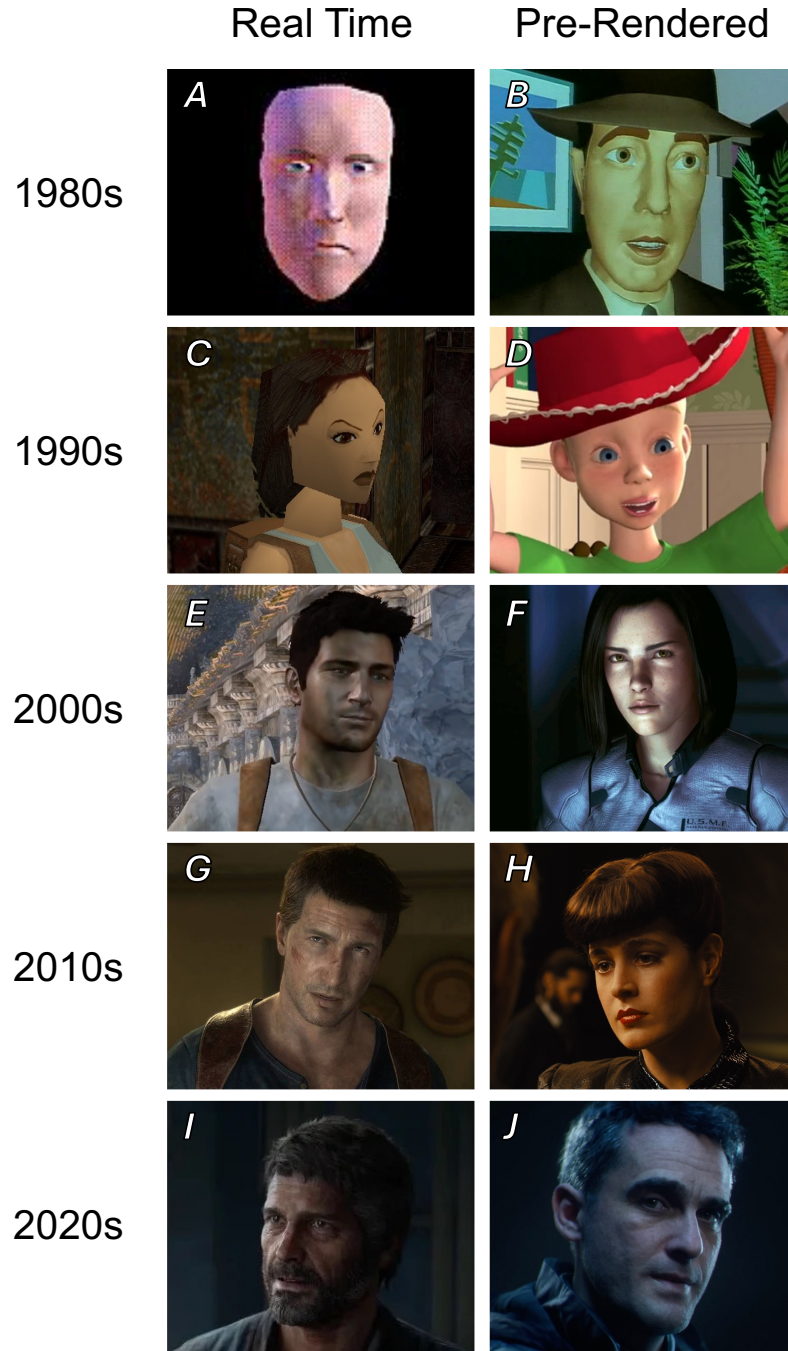


Figure 1.1: Examples of digital human characters across five decades, comparing real-time (RT) and pre-rendered (PR) rendering pipelines. Real-time characters are generated instantly by a graphics engine and are designed for interactive media such as video games, whereas pre-rendered characters are computed offline without real-time constraints, typically for film production. The figure illustrates the progression of digital human realism from early polygonal models to contemporary high-fidelity characters with physically based rendering and advanced facial animation, including (A) Mike the Talking Head (1988; RT), (B) Rendez-vous in Montreal (1987; PR), (C) Tomb Raider (1996; RT), (D) Toy Story (1995; PR), (E) Uncharted: Drake's Fortune (2007; RT), (F) Final Fantasy: The Spirits Within (2001; PR), (G) Uncharted: A Thief's End (2016; RT), (H) Blade Runner 2049 (2017; PR), (I) The Last of Us: Part I (2022; RT), and (J) Blue Dot (2024; PR).

understanding how people navigate social life.

Digital faces offer advantages for this kind of research. Unlike real human faces, digital stimuli can be precisely controlled and reproduced across participants. Researchers can systematically adjust specific features, such as face shape, skin tone, gaze direction, or patterns of facial movement, while holding other properties constant. This level of control makes it easier to test cause and effect relationships between facial cues and social judgments. Studying responses to digital faces also makes it possible to examine how people react to artificial agents compared with real humans. As digital avatars become more common in games, virtual reality, films, and AI chatbots, understanding these similarities and differences becomes increasingly important for both psychological theory and the practical design of believable digital characters.

1.2 Technological progress in rendering and animation

Because faces are processed with such perceptual precision and carry so much social meaning, efforts to make digital humans believable have required not only artistic creativity but also major technological innovation (Zhang et al., 2024). The challenge of meeting these perceptual expectations emerged gradually from early research into computer graphics. In the 1960s and 1970s, researchers were mainly concerned with how to represent three dimensional objects on a flat screen and decide which parts of those objects should be visible from a particular viewpoint (Clark, 1976). This process, often called visibility determination, ensures that only the surfaces that should be seen are drawn (Sutherland et al., 1974). At this stage, the goal was simply to display solid shapes correctly and consistently on screen.

As hardware improved in the 1980s and 1990s, graphics processing units, specialised processors designed to handle the large number of calculations required for rendering images, made more complex lighting and shading possible (Peddie, 2022). Programmable shaders, which are small programs that determine how surfaces respond to light, allowed artists to simulate reflections, shadows, and surface detail with greater flexibility (Bailey & Cunningham, 2009). However, despite these advances, surface realism relied heavily on artistic expertise rather than physical accuracy.

Techniques such as texture mapping and normal mapping further improved the appearance of digital surfaces (Heckbert, 1986). Texture mapping involves wrapping a flat image around a three-dimensional model, much like wrapping printed foil around a chocolate animal. The shape underneath does not change, but the printed layer gives it colour, patterns, and visible detail. Normal mapping adds the illusion of small bumps and grooves by adjusting how light reacts to the surface (Luhtapuro, 2014). It is similar to drawing shadows onto a smooth surface to make it appear textured, even though the underlying geometry remains flat. Together, these

techniques allowed digital objects to look richer and more detailed without increasing geometric complexity.

However, texture maps and normal maps are only part of how a surface appears. The final look also depends on its material and shader (Dempski & Viale, 2005). A material defines properties such as how shiny, rough, metallic, or transparent a surface should be. A shader is the program that calculates how light interacts with that material. For many years, these properties were adjusted largely by eye. Artists manually tuned reflectance and colour variation by comparing their digital models to photographs or real objects. Whether skin appeared soft, metal appeared polished, or fabric appeared worn depended heavily on the skill and judgement of the individual artist. For a long time, realism in digital surfaces was guided more by artistic interpretation than by consistent physical rules about how light behaves.

A major paradigm shift occurred with physically based rendering. Physically based rendering, or PBR, models how light interacts with surfaces according to physical laws (Pharr & Humphreys, 2010). It simulates properties such as energy conservation, meaning a surface cannot reflect more light than it receives, and light transport, which describes how light reflects and scatters through a scene. Techniques such as ray tracing and global illumination laid the foundations for these systems (Kajiya, 1986; Whitted, 1979). Ray tracing is a rendering method that simulates light by tracing the paths of rays as they travel from a camera into a scene, bounce off surfaces, and interact with materials. Global illumination refers to lighting that includes indirect light, meaning light that bounces off surfaces and softly fills a scene, rather than only direct light from a source.

Historically, these physically accurate lighting techniques were primarily used in offline or pre-rendered contexts because of their computational demands (Mileff & Dudra, 2022). In contrast, real time systems such as video games must generate each frame instantly, often at sixty frames per second, which traditionally required simplified approximations of lighting and shading. This distinction between rendering pipelines in terms of computational demands is reflected in the parallel advances in real time and pre-rendered graphics illustrated in Figure 1.

For digital humans, one especially important effect is subsurface scattering (*Fig. I.2*). Subsurface scattering refers to the way light enters a translucent material, travels through layers beneath the surface, interacts with skin tissue and blood, and then exits at a different point. This interaction occurs across the entire face and body, giving skin its soft glow and subtle colour variation. A particularly noticeable example occurs when strong light shines through thinner areas such as the ears, which can appear reddish because light passes through the tissue and interacts with blood before emerging again (*Fig. I.2C*). When subsurface scattering is not simulated, skin may appear flat or plasticky. When it is modelled accurately, it produces a more natural sense of depth and realism (Debevec et al., 2000; Jensen et al., 2001).

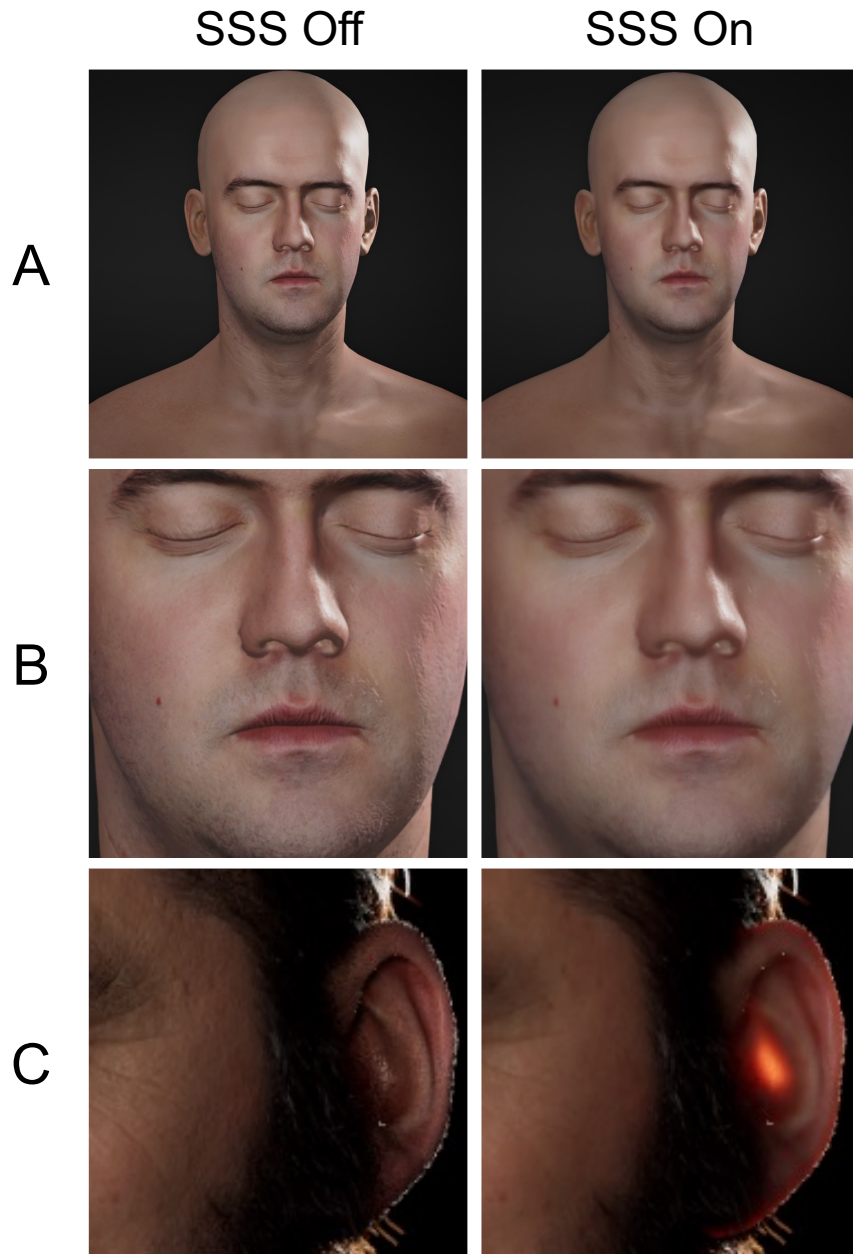


Figure 1.2: *Effect of subsurface scattering (SSS) on digital skin rendering. The left column shows rendering with subsurface scattering enabled, and the right column shows rendering with subsurface scattering disabled. (A) shows the full face, (B) shows a closer view of facial skin, and (C) shows a close-up of an ear. When SSS is enabled, light penetrates the skin, scatters beneath the surface, and exits at nearby points, producing softer shading, subtle colour variation, and a more natural sense of depth. This effect is particularly visible in thinner regions such as the ear, where transmitted light creates a reddish glow (Panel C). When SSS is disabled, skin appears more opaque, flatter, and more plasticky due to the absence of subsurface light transport.*

Rendering realism alone, however, is not sufficient for lifelike characters, as the structure and movement of the face are equally important. A facial rig is a structured control system that allows animators to move a digital face in a repeatable way (Hjelm, 2010). Early facial rigs were built using simple joint systems. Similar to the way a skeleton controls a body, a joint is a pivot point that allows parts of a model to rotate or bend. In early digital faces, a small number of joints might control the jaw, eyelids or eyebrows. While this allowed basic movement, it was difficult to produce subtle deformations such as cheek compression or fine wrinkles around the eyes. The resulting expressions often appeared mechanical because the face bent around rigid pivot points rather than deforming like soft tissue.

Modern systems often rely on blendshapes instead of joints alone (Shakir & Al-Azza, 2022). A blendshape is a pre modelled version of the face that represents a specific expression or muscle movement, such as a smile or a raised brow. The computer can smoothly interpolate, meaning gradually transition, between the neutral face and these predefined shapes. By combining multiple blendshapes at different strengths, complex and nuanced expressions can be produced. This approach allows the surface of the face to deform more like real skin and muscle, rather than rotating like connected pieces of rigid geometry.

Facial rigs are often organised according to the Facial Action Coding System, or FACS. FACS was developed by Ekman and Friesen (1978) as a scientific tool for describing facial movement in an objective and anatomically grounded way. Rather than labelling expressions with emotion words such as happy or angry, FACS breaks facial behaviour down into observable muscle actions (*Fig.1.3*). These basic components are called action units, which each correspond to a specific contraction or relaxation of one or more facial muscles, such as raising the inner brow or pulling the lip corners upward. This approach was designed to improve reliability in psychological research by allowing trained coders to describe facial behaviour in terms of visible movement rather than inferred mental states.

In computer graphics, FACS provides a structured and biologically meaningful framework for building facial rigs. Each action unit can be linked to a specific control or blendshape within the rig. This allows animators to activate muscle-based movements in a reproducible way. Using FACS also helps bridge research and production, because the same units used to analyse real human expressions can be used to generate them in digital characters.

The integration of high-resolution three-dimensional scanning, FACS based rigs, and physically accurate rendering has enabled digital humans with unprecedented visual fidelity. However, achieving this level of realism has traditionally been costly and labour intensive (Rocket-Brush Studio, 2025). Creating a photorealistic character often requires concept design, detailed modelling, high resolution textures and complex facial rigging capable of subtle muscle level deformation. Film and large-budget game productions frequently involve specialist teams and

Upper Face Action Units					
AU1	AU2	AU4	AU5	AU6	AU7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU41	*AU42	*AU43	AU44	AU45	AU46
					
Lip Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU9	AU10	AU11	AU12	AU13	AU14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU15	AU16	AU17	AU18	AU20	AU22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Pucker	Lip Stretcher	Lip Funneler
AU23	AU24	*AU25	*AU26	*AU27	AU28
					
Lip Tightener	Lip Pressor	Lips Parts	Jaw Drop	Mouth Stretch	Lip Suck

Figure 1.3: Overview of selected Facial Action Units (AUs) defined within the Facial Action Coding System (FACS). FACS decomposes facial behaviour into anatomically based components corresponding to specific muscle contractions or relaxations. The upper panel shows action units primarily involving the upper face, such as brow raising and eyelid movements, while the lower panel shows action units involving the lower face, including lip, jaw, and chin movements. Each action unit represents a distinct, observable facial movement. Reproduced from De la Torre et al. (2015) with permission © [2015] IEEE.

large budgets to build even a single high-quality character (Alexander et al., 2009). For many research teams, such resources are difficult to access, limiting the use of highly realistic digital humans in experimental settings.

Platforms such as Unreal Engine’s MetaHumans have begun to change this landscape. MetaHumans combine high resolution geometry, physically based skin and eye materials, advanced hair simulation, and detailed facial rigs within an openly accessible pipeline (Games, 2021), see *Fig. 1.4*. At the time of writing, MetaHumans represent one of the most accessible state-of-the-art digital human systems available without a paid licence, offering industry-level visual fidelity within a freely available real-time engine; for a broader overview of contemporary digital human creation tools see Mellberg (2025). By providing state of the art digital humans without the need for custom production pipelines, these tools reduce the financial and technical barriers that previously restricted photorealistic characters to large studios. What once required large, specialist teams and substantial budgets can now be achieved within research environments using widely available tools.

Although advances in rendering and animation have brought digital humans closer to photorealism than ever before, these developments have largely been driven by the entertainment industry rather than psychological research. Producing high-fidelity digital humans requires expertise spanning computer graphics, character modelling, facial rigging, animation, rendering, and software engineering, together with substantial financial investment and large multidisciplinary teams. Such resources are typically available at major game studios and film production companies but are rarely accessible to psychology research laboratories, which typically operate with considerably smaller teams and budgets. Therefore, the digital avatars used in psychological research have generally not kept pace with the technological advances seen in commercial entertainment, despite the growing importance of realistic virtual humans for studying social perception and interaction (Jack & Schyns, 2017).

However, making high-fidelity digital humans such as MetaHumans widely accessible does not by itself ensure that they are suitable for psychological research. Unlike entertainment applications, where realism is judged predominantly by the perception of the audience, experimental research requires a known ground truth so that facial behaviours can be manipulated systematically and interpreted consistently across studies. Epic Games state that the MetaHuman facial rig was developed by building upon the foundations of the Facial Action Coding System (FACS; Ekman and Friesen, 1978), with their technology providing a real-time implementation of FACS-based facial rigging (Epic Games, 2021). Although the MetaHuman facial rig does not use exactly the same naming conventions as FACS, all major FACS Action Units have corresponding controls within the rig, together with numerous additional controls that allow finer adjustment of facial expressions. However, while the system is rooted in FACS, the exact technological implementation of how FACS-based movements are translated into the MetaHuman

rig is not fully transparent. Consequently, visual realism and technical design alone are insufficient to establish scientific validity. Validation studies, such as the work presented in *Chapter 3*, therefore represent an important step towards the wider adoption of MetaHumans as experimental stimuli within psychological research.

Empirical work suggests that such improvements in character fidelity meaningfully affect perception. Higgins et al. (2021) examined different levels of detail within MetaHumans. A level of detail system refers to the use of multiple versions of the same character at different geometric complexities, allowing visual quality to be reduced when necessary for performance. They found that higher fidelity versions of the same character were judged as more humanlike and less eerie than lower fidelity versions, even when the animation was identical. Observers also perceived emotional expressions as more intense in the higher fidelity models. This finding indicates that visual realism can alter how motion is interpreted. Perception depends on the coherence between appearance and behaviour rather than on appearance alone.

These technological developments have brought digital humans closer to photorealism than ever before. Yet as realism increases, perceptual expectations also rise. When visual detail suggests biological plausibility, even small deviations in motion, timing, or coordination can become highly noticeable.



Figure 1.4: *Example of a MetaHuman rendered in Unreal Engine using physically based materials and real-time lighting. The image illustrates high-resolution facial geometry, detailed skin texture with subsurface scattering, realistic eye shading, and strand-based hair simulation. Platforms such as MetaHuman enable industry-level visual fidelity within an openly accessible real-time engine, reducing the technical and financial barriers to creating photorealistic digital humans for research and production contexts.*

1.3 The uncanny valley

Despite the advances in digital graphics solutions, evidence suggests that some computer-generated faces do not always engage face processing in the same way as real photographs. For example, Crookes et al. (2015) found that recognition memory was poorer for computer-generated faces than for real faces, with the largest drop in performance for own race faces. They also reported that a classic marker of face expertise, the other race effect, was substantially reduced when participants viewed computer-generated faces. The other race effect refers to the common finding

that people recognise faces from their own racial group more accurately than faces from other groups. When this effect is weaker for computer generated faces, it suggests that these stimuli may not fully recruit the specialised expertise people typically use for recognising real faces. This limitation is important for experimental design, and it may also help explain why some digital humans feel subtly less lifelike.

This unease is often discussed using the concept of the uncanny valley. The uncanny valley was first proposed by the roboticist Masahiro Mori (1970), later translated into English by Mori et al. (2012). Mori suggested that as an artificial agent becomes more humanlike, people tend to feel more positive towards it, but only up to a point (*Fig.1.5*). When the agent becomes almost human, but still contains small imperfections, people may instead experience discomfort or eeriness. Mori described this as a sharp dip in emotional response, which he called the uncanny valley. He used the example of a prosthetic hand, which may look human at a glance, but it can feel cold or unnatural when touched, creating an unsettling sense of mismatch.

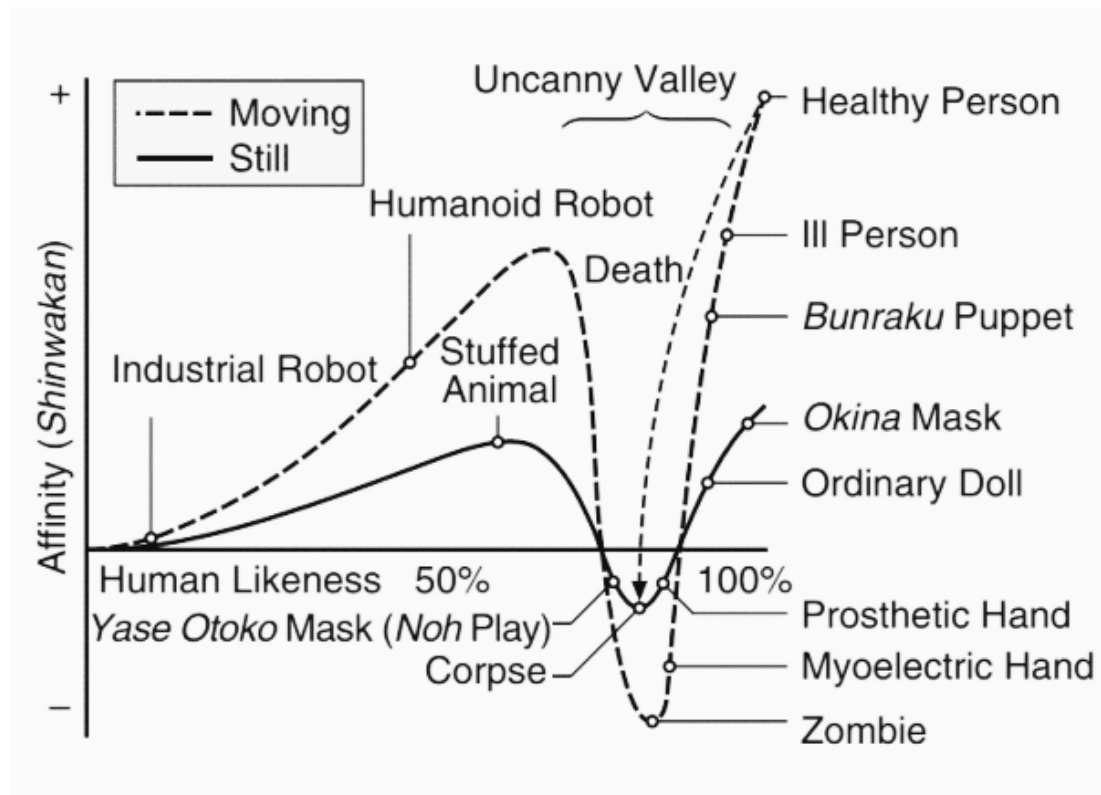


Figure 1.5: Original illustration of the uncanny valley hypothesis proposed by Mori (1970). The vertical axis represents affinity, and the horizontal axis represents increasing human likeness. As artificial agents become more humanlike, affinity increases until a critical point at which near-human entities elicit a sharp decline in emotional response, forming the so-called uncanny valley. The dashed line indicates that movement can intensify this effect, producing a deeper valley compared with still entities. Examples such as prosthetic hands, corpses, and zombies are positioned within the valley, whereas healthy humans occupy the highest level of affinity. Adapted with permission from Mori et al. (2012); © 2012 IEEE.

Mori also suggested that movement can intensify the uncanny valley effect. A prosthetic limb that moves in an unnatural way may feel more disturbing than one that remains still. This idea has become a central theme in later research. The uncanny valley is now widely understood not as a simple response to realism, but as a response to inconsistency. When different cues suggest different levels of humanness, the character may feel unsettling.

A major review by Kätsyri et al. (2015) evaluated empirical evidence across uncanny valley studies and concluded that the valley is not inevitable. They argued that the original formulation can be interpreted in multiple ways, and that the naïve claim that increasing human likeness automatically produces a drop in affinity is not well supported. Across studies, increasing human likeness was often associated with more positive evaluations rather than a sharp decline at near-human levels (Cheetham et al., 2014; Looser & Wheatley, 2010; Seyama & Nagayama, 2007). The authors therefore suggested that the uncanny valley appears only under specific conditions.

One explanation that received little consistent support was the categorisation difficulty account, which proposes that negative reactions arise because observers struggle to classify a character as either human or artificial (MacDorman & Ishiguro, 2006; Pollick, 2010; Ramey, 2005). Although several studies demonstrated categorical perception along human–artificial continua (Cheetham et al., 2011; Looser & Wheatley, 2010), evidence linking category ambiguity directly to increased eeriness was weak (Cheetham et al., 2014). Instead, the strongest empirical support was found for perceptual mismatch accounts, which propose that negative reactions arise when different cues to humanness are inconsistent with one another.

For example, characters with highly realistic eyes paired with artificial faces, or artificial eyes embedded in otherwise humanlike faces, were rated as more eerie than characters whose features were consistently stylised or consistently realistic (MacDorman et al., 2009; Seyama & Nagayama, 2007). Similar effects were observed when realistic human faces were paired with synthetic voices (Mitchell et al., 2011). The review also highlighted evidence that observers are particularly sensitive to atypical features on otherwise humanlike faces, such as exaggerated eye size (MacDorman et al., 2009). Kätsyri et al. (2015) concluded that the uncanny valley is best understood as emerging from specific forms of perceptual mismatch, rather than from human likeness alone.

A later meta-analysis by Diel et al. (2021) examined seventy-two experimental studies that explicitly tested for an uncanny valley effect. Across 247 effect sizes, they reported a large overall effect, indicating that negative reactions to highly humanlike artificial entities are reliable across many paradigms. Importantly, they also found substantial variation in methods, including ten different stimulus creation techniques and more than fifty different dependent measures. While this diversity highlights a lack of methodological consensus across the field, it also highlights the robustness of the phenomenon.

When the authors compared different stimulus creation techniques, they found that face distortion produced the largest effect sizes. Face distortion refers to directly altering the proportions or spatial arrangement of facial features, such as changing the distance between or enlarging the eyes, or shifting the position of the mouth. Studies that manipulated these facial proportions, including MacDorman et al. (2009) and Mäkräinen et al. (2014), consistently produced strong feelings of eeriness. This sensitivity is often described as configural processing, which refers to perceiving a face as an integrated whole defined by the spatial relationships between features, rather than as a collection of independent parts.

Large effects were also observed for other techniques, including distinct entities, which compare clearly different categories such as robots and humans; realism rendering, which increases visual detail and surface realism; and morphing, which blends one face into another along a continuum. However, the authors noted that morphing between very similar endpoint stimuli did not always produce a clear valley pattern (Cheetham et al., 2015; Kätsyri et al., 2019). They also found that studies using non-human animal stimuli could produce strong uncanny effects, suggesting that violations of expected biological norms may be more important than human likeness alone (Schwind et al., 2018; Yamada et al., 2013). These findings support the idea that deviations from typical structural patterns, especially in the spatial configuration of facial features, may play a central role in eliciting eeriness (Chattopadhyay & MacDorman, 2016; Diel & MacDorman, 2021).

Diel et al. (2021) also analysed how the uncanny valley effect was measured across studies. They found that the effect size was larger on the ‘human side’ of the valley than on the ‘non-human side’. This means that emotional reactions changed more strongly as artificial characters became highly humanlike, compared with when they were clearly mechanical or stylised. In other words, the sharpest shifts in evaluation tended to occur close to human realism rather than far from it.

The strength of the effect also depended on how it was measured. Studies that used direct measures of eeriness or uncanniness, such as ratings of eerie or creepy, produced larger and more consistent effects than studies that relied on broad positive affect measures such as likable or pleasant (C.-C. Ho & MacDorman, 2010, 2017; Kätsyri et al., 2017). Indirect measures, including categorisation reaction time or behavioural avoidance, showed more variable results (Cheetham et al., 2013; Strait et al., 2015). On this basis, clearly defined, uncanny valley specific affect measures and carefully controlled stimulus manipulations should be used to study the phenomenon. Overall, these findings support the existence of the uncanny valley effect while highlighting the importance of precise definitions, consistent measurement, and stronger experimental control.

The uncanny valley has also been described using more specific emotional dimensions. Ho

and MacDorman (2010) argued that perceived humanness and eeriness should be treated as separate dimensions, rather than opposite ends of a single scale. Humanness refers to how human an entity appears. Eeriness refers to an unpleasant emotional response characterised by discomfort, fear, creepiness, or unease. Ho and MacDorman later refined these measures and showed that they can capture patterns consistent with Mori's theory, where a character may be rated as highly humanlike while still producing high eeriness (Ho & MacDorman, 2017). This approach is important because it allows researchers to test uncanny valley predictions more directly than general measures such as likeability.

A key implication of this work is that the uncanny valley is related to realism and whether the cues of a character form a coherent whole. Historically, many digital humans were built using pipelines where geometry, textures, and motion were treated separately. However, human perception integrates faces holistically. This means that even if each part of the character is strong in isolation, the overall result can still feel wrong if the parts do not match.

1.4 Motion and realism

In Mori's original description of the uncanny valley, movement was expected to make the dip in comfort stronger, because motion makes it harder to ignore small errors in how an agent behaves (Mori, 1970; Mori et al., 2012). Later evidence fits this general idea, but it also shows why motion has been difficult to study cleanly. Negative reactions tend to appear under specific conditions, especially when different cues to humanness do not match each other (Kätsyri et al., 2015). In practice, many uncanny valley studies still rely heavily on static images or treat motion as a secondary factor, which helps explain why findings about motion have often been mixed across the literature (Diel & Lewis, 2022; Kätsyri et al., 2015).

Even so, several studies show that motion and appearance can interact in a way that matches Mori's intuition. McDonnell et al. (2012) held the underlying performance constant by using the same motion capture animations across multiple rendering styles. They found that highly realistic and highly abstract styles were often judged more appealing than styles in the middle range, and the least appealing character became even less appealing when it was shown moving rather than still. They also tested what they called motion anomalies, meaning deliberately introduced animation errors. In their case, the strongest anomaly removed motion from one half of the face, creating an obvious breakdown in facial coordination, while a weaker anomaly used static eyes, meaning the eyeballs did not rotate while the rest of the face moved. When these large motion anomalies were present, characters with photographic human skin textures were judged more unpleasant than more stylised characters, which supports the idea that a more realistic look can raise expectations for realistic movement and reduce tolerance for visible errors.

A key limitation of the broader motion literature is that it has mainly focused on expressive or goal directed behaviour, such as emotional facial expressions and speech related facial movement, because these are easier to define, easier to label, and easier to validate against performance capture or acting reference. By comparison, idle motion has received far less attention. Idle motion refers to the small, ongoing, low intensity movements that occur when a person is not speaking or acting, but is still visibly alive, such as natural blinking, tiny gaze shifts, subtle head and torso adjustments, and small mouth and facial muscle changes that do not form a clear expression. Idle motion matters because it is the behavioural baseline for many digital humans in games, film close ups, virtual reality, and conversational agents. Viewers spend a lot of time looking at characters who are waiting, listening, or simply present on screen. If this baseline behaviour is missing, overly regular, or inconsistent with a realistic face, the character can feel stiff, puppet like, or subtly wrong, even when nothing dramatic is happening (Diel & Lewis, 2022; Kättyri et al., 2015).

1.5 Modelling human perception

To determine whether a particular visual or motion cue influences perception, that cue must vary (Jhangiani, 2018). If the features of a stimulus remain constant, it is not possible to identify which aspects are responsible for differences in how it is judged. For example, if a face is always shown with the same eyebrow shape, it is not possible to know whether that eyebrow shape influences perception of the face. Only by systematically changing the shape while holding other features constant can its effect be examined. By introducing controlled variations in a stimulus and measuring how observers respond, researchers can examine how specific changes relate to differences in perception. Psychophysics provides a framework for this approach. Psychophysics is the study of the relationship between measurable properties of stimuli and subjective experience (Gescheider, 2013). It investigates how changes in the physical characteristics of a stimulus correspond to changes in what people perceive or report.

One widely used psychophysical method is reverse correlation (Brinkman et al., 2017; Jack & Schyns, 2017). Reverse correlation is a data-driven technique that estimates which stimulus features influence judgments by relating random stimulus variation to patterns in observer responses (*Fig. 1.6*). In a typical study, participants are shown many slightly different versions of the same neutral face. On each trial, random visual noise is overlaid on this neutral face image. For instance, the noise may slightly lighten or darken certain regions or subtly alter shading around the mouth or eyes. Participants may be shown two noisy versions of the same neutral face per trial and asked which one looks happier, or which one looks sadder (Jack & Schyns, 2017).

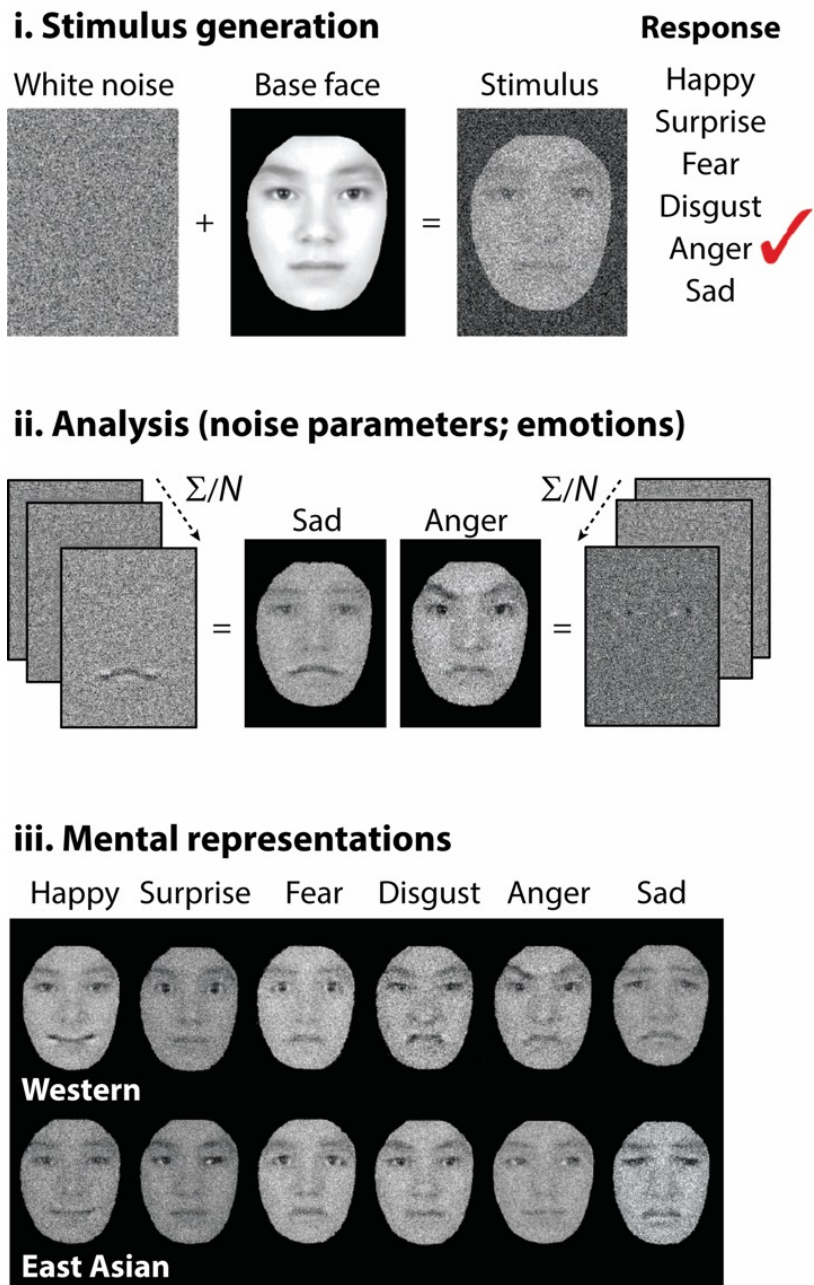


Figure 1.6: Illustration of the reverse correlation method applied to facial perception. In the stimulus generation stage (i), random white noise is added to a neutral base face to create slightly varied stimulus images. Participants categorise each noisy face according to emotional judgments (e.g., happy, anger, sad). In the analysis stage (ii), noise patterns associated with each response category are averaged across trials to estimate the visual features influencing those judgments. The resulting classification images represent observers' mental representations of different emotions (iii), shown here for Western and East Asian participants. Adapted from Jack & Schyns (2017) with permission.

Importantly, the underlying face is always the same as the only difference between the two images is the random noise pattern added on top. Over many trials, the researcher records which noisy versions are judged as happier and which are judged as sadder. By averaging the noise patterns associated with each response, the experimenter can reconstruct the visual features that influenced the judgment. This averaged pattern is known as a classification image, and it provides an estimate of the perceptual template guiding the decision (Eckstein & Ahumada, 2002). For example, if noise that slightly darkens the corners of the mouth is consistently associated with sad faces, that pattern will emerge in the reconstructed image. In this way, the method reveals the visual cues observers rely on, even though those cues were not explicitly designed into the stimulus. The technique was first developed in auditory psychophysics by Ahumada and Lovell (1971), where random noise was added to sound signals, and was later adapted to vision and face perception.

Reverse correlation has been widely used in social perception. For example, Dotsch and Todorov (2012) used reverse correlation to reveal the facial features that drive judgments of trustworthiness and dominance. More broadly, researchers have used the method to study mental representations of emotions and stereotypes (Brinkman et al., 2017). Reverse correlation is valuable because it does not require strong assumptions about which cues matter. Rather, it allows the observer's own perceptual template to emerge from the data.

Reverse correlation has also been extended to dynamic stimuli. Rather than applying pixel noise to static images, researchers can generate large sets of animated faces by systematically varying facial action unit patterns over time. Yu et al. (2012) introduced the Generative Face Grammar (GFG), a parametric face modelling system in which dynamic facial expressions are constructed from random combinations of action units, each controlled by a parameterised motion curve. This framework allowed facial structure and movement to be manipulated in a systematic and reproducible way, making it possible to apply reverse correlation methods to dynamic facial behaviour. By relating randomly varied motion patterns to observer judgments, models could be derived identifying which facial movements and temporal dynamics communicate particular emotions. This work demonstrated that reverse correlation can be applied not only to static appearance, but also to temporal facial motion. While the GFG provided a powerful and methodologically influential framework for dynamic reverse correlation research, it was developed within the technological constraints of its time. Another tool often used in this field is FaceGen (Banz & Vetter, 1999; Oosterhof & Todorov, 2008; Todorov et al., 2025) which, like the GFG, is based on a statistical three-dimensional morphable model derived from scanned human heads. FaceGen allows systematic manipulation of facial shape and texture and has been used in studies of social face perception. However, compared with the GFG (*Fig. 1.7*), it provides more limited control over fine-grained facial action unit dynamics, and neither system reflects the level of geometric detail and physically based rendering now possible in modern real-time

engines.

Overall, reverse correlation is now a well-established method in vision science and social perception. In most applications, the focus has been on clearly interpretable signals such as smiling or frowning, or other overt displays of mental state. Subtler forms of facial behaviour, especially low intensity movements that occur outside clear emotional expressions, have received comparatively less attention. This leaves important questions about how such behaviour contributes to impressions of naturalness and aliveness, and more specifically how subtle facial motion shapes the perception of digital avatars.

Recent advances in generative artificial intelligence have led to the development of diffusion-based approaches for generating realistic facial animation and other forms of dynamic visual content (Blattmann et al., 2023; Ho et al., 2022; Miao et al., 2025; Rombach et al., 2022). Rather than relying on manually defined animation parameters, these methods generate motion from large datasets and are capable of producing highly realistic and temporally consistent animations (Blattmann et al., 2023; Miao et al., 2025). However, the latent representations learned from the datasets used by diffusion models do not correspond to explicit or independently controllable facial behaviours (Rombach et al., 2022). Thus, it is difficult to systematically manipulate individual motion characteristics or directly relate perceptual responses to specific aspects of facial movement. A recent survey identifies controllability and interpretable control as an ongoing challenge for diffusion-based animation systems, despite substantial improvements in realism (Miao et al., 2025). The reverse correlation methodology adopted throughout the present thesis required animations to be generated from explicit and consistent parameters that can be manipulated, allowing participant judgements to be mapped directly onto the facial movements used to generate each animation. Therefore, a fully parametrised animation framework was developed and described in *Chapter 2*, providing the level of experimental control required for the studies presented in this thesis.



Figure 1.7: *Example of a facial stimulus generated using the Generative Face Grammar (GFG; Yu et al., 2012). Although highly controllable and methodologically influential in reverse correlation research, the visual fidelity reflects the graphics standards available at the time of its development.*

1.6 The present thesis

The present thesis addresses this gap. It brings together computer graphics, psychophysics, and social cognition to investigate how subtle idle facial motion contributes to the perception of aliveness in digital avatars. A fundamental component of this work is the development of a scalable animation framework for generating large numbers of high fidelity MetaHuman stimuli suitable for reverse correlation research. In addition, the thesis benchmarks MetaHumans

against GFG avatars. The GFG provided an important parametric tool for studying facial perception and was pivotal for reverse correlation research at the time of its development (around 2010; Garrod et al., 2010). However, rapid advances in computer graphics technology over the past decade and a half mean that these stimuli are now dated by contemporary visual standards. While methodologically influential, they do not reflect the level of geometric detail and animation complexity now possible in modern graphics engines. High fidelity MetaHumans in Unreal Engine provide physically based skin rendering and sophisticated facial rigs (Epic Games, 2025; Games, 2021). Because these rigs are compatible with action unit style control, subtle facial movements can be manipulated precisely. This allows the creation of stimuli that are both experimentally controllable and state-of-the-art level of current computer graphics.

Using these stimuli, the thesis applies reverse correlation methods to estimate perceptual templates of natural idle facial motion. In this context, the category of interest is not an emotion or a static trait, but the impression that a face looks alive and natural when doing very little. Participants are presented with large numbers of short idle animations and provide perceptual judgments. These responses are then correlated with the underlying motion parameters used to generate the animations. This makes it possible to infer which patterns of idle facial motion are most strongly associated with perceived naturalness.

This approach has practical advantages. It does not assume which micro-movements matter. Instead, it allows perceptually important cues to emerge from the data. It also provides a direct bridge between perception and animation. Rather than simply describing which animations are preferred, the method produces models that can be used to generate new motion profiles. This makes it possible to test causal predictions about how motion affects social evaluation.

The implications are particularly relevant to the uncanny valley. By clarifying how subtle idle motion shapes perceived aliveness, this work helps explain when highly realistic digital characters feel coherent and lifelike, and when mismatches between appearance and motion may produce discomfort or eeriness. Understanding these perceptual mechanisms is essential for identifying how and why otherwise advanced digital humans can become unsettling.

In summary, digital humans have progressed from rigid and stylised models to highly realistic avatars capable of subtle facial motion. Research on the uncanny valley shows that realism can backfire when cues do not match, especially when appearance and motion conflict (Diel et al., 2021; Kätsyri et al., 2015; Mori, 1970; Mori et al., 2012). Motion, particularly idle motion, is central to believability (McDonnell et al., 2012). At the same time, psychophysical approaches such as reverse correlation provide a powerful way to uncover the perceptual cues that drive social judgments (Ahumada & Lovell, 1971; Dotsch & Todorov, 2012; Jack & Schyns, 2017; Yu et al., 2012). The present thesis combines these developments to model how subtle facial motion contributes to perceived aliveness in modern digital avatars.

Chapter 2 “Unreal Engine Animation Framework”: introduces the Unreal Engine based animation framework developed for this thesis. This chapter explains how MetaHuman characters were adapted into a system that can generate large numbers of high-fidelity animation stimuli with precise and reliable control over facial motion. It establishes the technical foundation needed for the empirical work that follows and motivates why automation and standardisation are essential for perception experiments that require hundreds or thousands of trials.

Chapter 3 “Validating MetaHuman Avatars: A Mixed Methods Benchmark”: presents the first two empirical studies, which validate MetaHumans as psychological research stimuli. The studies in this chapter compare MetaHumans with Generative Face Grammar avatars, using the same characters and matched video conditions, and measures participants’ impressions of humanness, eeriness, realism, and familiarity (i.e., uncanny valley related measures) along with more subjective evaluations. This chapter provides the first evidence that MetaHumans offer practical advantages for perception research, and justifies their use as the stimulus platform for the later reverse correlation experiments.

Chapter 4 “Reverse Correlation of Natural Idle Facial Motion”: reports the main reverse correlation study, which forms the central empirical contribution of the thesis. In this study, participants rated large sets of procedurally generated idle facial animations, allowing facial action units and motion parameters to be related directly to perceived naturalness. This data was used to build perception driven models of idle facial motion, identifying which movements and timing patterns are associated with more natural impressions. These models are then validated in a subsequent stage of the same study, in which new animations synthesised from low, medium, and high naturalness motion profiles are presented to new observers to test whether the intended differences are reflected in their ratings. Specifically, the validation tests whether high model animations are judged as more natural than medium, and medium more natural than low, demonstrating the predictive and generative validity of the models.

Chapter 5 “Motion Naturalness and the Uncanny Valley in Digital Avatars”: uses the validated motion profiles from Chapter 4 to test a theoretical question about the uncanny valley. This study holds visual appearance constant while systematically manipulating idle motion naturalness, allowing motion to be isolated as an independent causal factor in social evaluation. It examines whether unnatural idle motion increases eeriness and reduces humanness, realism, and familiarity, providing a direct test of mismatch-based accounts of the uncanny valley within a modern, high-fidelity digital human platform.

Chapter 6 “General Discussion”: brings together the main findings of the thesis and considers their broader theoretical and practical implications. It reflects on what the results suggest about digital human realism and social perception, discusses key limitations, and outlines directions for future research.

Chapter 2

Unreal Engine Animation Framework

2.1 Chapter Overview

This chapter introduces the technical foundation that underpins all subsequent experimental work in this thesis: a custom Unreal Engine-based animation framework for generating high fidelity, parametrically controlled digital human stimuli. Building on the rationale established in Chapter 1, the chapter focuses on the technical implementation of a contemporary, photorealistic character solution using MetaHumans for reverse correlation studies. Subsequently, the chapter describes the architecture of the proposed pipeline, from Python-based animation curve generation and CSV data structures through to Unreal Engine integration via Blueprint systems for automated stimulus recording and rendering. Practical considerations such as performance constraints, storage demands, and scalability for large-scale experiments are discussed. The chapter concludes by positioning this framework as a MetaHuman based equivalent of the GFG pipeline for dynamic face stimuli for reverse correlation, and outlining how it will be used in the following empirical chapters of this thesis.

2.2 Motivation and Requirements

Research on face perception and social judgement has been strongly influenced by the development of parametrically controlled, 3D face stimulus frameworks that allow the generation of large and systematically varied sets of face stimuli with high experimental precision. One of the most influential approaches involved combining the Generative Face Grammar (GFG) face stimuli generation pipeline with social psychophysics approaches to map perceptual judgements onto underlying facial features and dynamics with high control and reproducibility (Hensel et al., 2020; Jack & Schyns, 2017; Yu et al., 2012). Parallel work in social evaluation used similar

parametric approaches to investigate how face topology and texture drive the perceptions of various traits such as trustworthiness and dominance, or extroversion and introversion (Oosterhof & Todorov, 2008; Todorov & Oosterhof, 2011). Collectively, these tools and approaches established parameterised 3D face frameworks as powerful methodological platforms for probing underlying perceptual mechanisms.

One approach for probing such perceptual mechanisms is a social psychophysics technique known as reverse correlation (RC). RC is a data-driven technique for estimating the perceptual representations that guide human judgements. In its simplest form, RC works by systematically sampling feature variations across many stimulus instances, collecting behavioural responses to those variations, and statistically grouping the resulting response-stimulus relationships to reveal which features most strongly influence perception. Early implementations of RC applied this logic by varying pixel-level noise over a fixed 2D base face (Dotsch & Todorov, 2012; Gosselin & Schyns, 2001, 2003; Mangini & Biederman, 2004; Murray, 2011), enabling researchers to infer which regions of an image were most influential for particular judgements.

Subsequent developments extended the method to fully parameterised 3D faces and dynamic expressions, making it possible to identify the specific facial features and motion parameters that carry perceptual weight for emotion, identity, and social traits (Brinkman et al., 2017; Jack & Schyns, 2017; Jack et al., 2014; Richoz et al., 2015; Todorov & Oosterhof, 2011). In such dynamic 3D face systems, facial motion is typically organised into action units (AUs), which represent specific muscle-based movements such as raising the brow or pulling the lip corners (Ekman & Friesen, 1978). Each AU can be controlled numerically with an activation curve that varies over time; the AU activation curve can be controlled with parameters such as how strong the activation is, and when it begins and ends. In RC paradigms, these activation curve parameters can be systematically varied and statistically linked to participants' judgements. This requires an animation framework in which facial movement is precisely controllable and reproducible. RC typically requires from hundreds to thousands of systematically varied stimuli per experiment in order to reliably estimate perceptual representations (Brinkman et al., 2017; Brinkman et al., 2019; Jack, Garrod, et al., 2012). Generating such large stimulus sets manually is not feasible, making automated and scalable animation pipelines a methodological necessity rather than a convenience. As the applications of RC for face perception have expanded, so has the need for stimulus generation frameworks that retain precise parametric control while meeting contemporary expectations for digital human face rendering in terms of animation quality and 3D model realism.

As discussed in *Chapter 1*, recent advances in digital human technology have substantially improved the visual and animation fidelity of avatar systems, while research has shown that higher fidelity characters can influence perceptions of humanness, eeriness, and emotional expression (Higgins et al., 2021). Therefore, the present framework adopts Epic Games' MetaHu-

man ecosystem as a contemporary stimulus platform. Rather than reiterating the broader rationale for adopting MetaHumans established in *Chapter 1*, the remainder of this chapter focuses on the specific features and technical implementation that enable a parametrically controlled animation pipeline suitable for reverse correlation experiments.

The MetaHuman ecosystem directly addresses these requirements by providing fully rigged, photorealistic digital humans designed for detailed facial and full-body animation within Unreal Engine. Out of the box, among other features, the MetaHumans offer: consistent and detailed 3D character topology with high polygon counts, high-quality physically-based skin, eye, and hair materials, and a comprehensive character Control Rig that supports keyframed, procedural, and motion capture animation. A key component of this realism lies in the geometric fidelity of the facial mesh itself. Higher polygon density and well-structured topology allow smoother deformation, particularly around the eyes and mouth, supporting subtle facial movements. In lower fidelity meshes, limited resolution can introduce artefacts and reduce motion precision, which becomes a methodological limitation when studying fine-grained facial dynamics. Both the facial and full body rig systems enable precise manipulation of facial expressions, gaze, head motion, and body posture. For researchers, this combination effectively bridges state-of-the-art realism with fine grained parametric control, substantially reducing engineering overhead while allowing experimental manipulations to remain precise, replicable, and consistent across a variety of MetaHuman characters (see MetaHuman official documentation for more technical details; Epic Games, 2025).

With Unreal Engine 5.6, the MetaHuman ecosystem has developed into a unified, in-editor solution for creating, animating, and rendering high fidelity digital humans. This integration allows researchers to generate and adjust characters, facial expressions, and full-body movements within a single environment, greatly expanding the possibilities and allowing for more streamlined workflows. The introduction of parametric bodies extends MetaHuman’s detailed face creation controls to fully customisable body proportions, enabling systematic manipulation of body morphology alongside facial structure. This represents a significant advance over most existing 3D character frameworks used in psychophysical experiments to date, which typically provide only head and face models without fully articulated bodies. With MetaHumans, researchers can investigate how facial motion interacts with gaze, head movement, posture, and whole-body cues within a single, consistent rig, broadening the scope of perceptual studies beyond facial features alone and into more multi-faceted human behaviour.

One of the major advantages of the MetaHuman framework is that it can be used across many different software tools and pipelines while keeping the same standardised character structure. Tools such as Mesh to MetaHuman allow existing character meshes or scans to be transferred onto the standard MetaHuman rig, making it possible to integrate legacy assets or create bespoke avatars without compromising facial fidelity or rig consistency. This standardisation en-

tures that deformation, animation, and control behave predictably across characters which is an essential requirement for psychophysical and RC paradigms where stimuli must be consistently manipulated. The updated licensing model in Unreal Engine 5.6 further enables MetaHumans to be used across different engines and software pipelines, reducing barriers for long-term reproducibility and collaboration across labs.

Other areas of psychological and applied research are already beginning to adopt MetaHumans as a high-fidelity stimulus platform. MetaHuman characters have been used to study photorealism and empathy (Higgins et al., 2021, 2023), avatar realism in social interaction (Fraser et al., 2024), uncanny valley and user acceptance effects (Bae et al., 2025), digital doubles (Liu et al., 2025), and even interactive educational simulations (Jang et al., 2025). These examples illustrate the increasing use of MetaHumans for empirical research.

Although Unreal Engine is first and foremost a commercial video-game engine, its power and flexibility have led to adoption across many industries including film and virtual production, digital twins and simulation, architectural visualisation, automotive design, virtual training, and immersive XR applications (Chia, 2022; Jungherr & Schlarb, 2022). However, this breadth also means that Unreal Engine is not designed out of the box for generating high volume, precisely controlled psychophysical stimuli. Researchers must therefore add an additional layer of software engineering to adapt its production-oriented tools into the kind of parametrically precise, reproducible pipelines required for experimental work.

The present chapter introduces such adaptation in the form of a custom animation framework developed for this thesis that connects externally generated parametric motion data to the MetaHuman Control Rig in Unreal Engine for automating stimuli generation. This system was explicitly designed to mirror the functional principles of the Generative Face Grammar (GFG; Yu et al., 2012) pipeline but implemented with state-of-the-art MetaHuman characters and the modern animation infrastructure of Unreal Engine. The workflow allows procedurally generated motion curves, to be applied directly to MetaHumans and rendered as tightly controlled, high-fidelity experimental stimuli at scale. Although built for reverse-correlation experiments on facial motion, the underlying approach can generalise to any research domain requiring precise, reproducible control over photorealistic animated digital humans.

2.3 Architecture Overview

This section outlines the structure and implementation of the procedural animation pipeline developed to generate MetaHuman RC stimuli within Unreal Engine. The pipeline allows parameterised motion data created in Python (or other software) to be applied to MetaHuman characters via .csv data import, Control Rig mapping (Unreal Engine’s built-in system for an-

imating characters directly in the engine by driving rig controls and custom parameters), and automated recording with the Take Recorder (a tool for capturing real-time character animations into reusable sequence assets). The framework was designed to provide precise experimental control, reproducibility, and scalability for high-fidelity behavioural research.

The motion data applied within this workflow is generated externally using reverse-correlation sampling principles, following the parametric logic originally formalised in 3D frameworks such as the GFG (Yu et al., 2012). A detailed description of the curve generation process used in this thesis is provided in Chapter 5 where we conduct a reverse correlation experiment with the animation framework described in the present chapter.

2.4 CSV Schema

At the core of this animation workflow lies a comma-separated value (.csv) file, which defines the temporal structure of every animation and provides the link between the Python-based animation curve generation and Unreal Engine’s animation system. The file serves as an engine-independent definition of motion, specifying how each facial, head, or eye control changes frame by frame throughout an animation sequence.

Each row in the file corresponds to a single frame of animation, while each column corresponds to a control channel (such as a facial Action Unit activation) or a piece of metadata describing that frame. In its simplest form, the file begins with two key columns: VideoNumber and FrameNumber, followed by one or more control columns representing the active animation parameters.

- **VideoNumber [integer]:** indicates which video clip a given frame belongs to. This allows multiple animations to be stored within a single .csv file, reducing data file volume.
- **FrameNumber [integer]:** indexes each frame within a clip, starting from 0 up to $n - 1$, where n is the total frame count of the animation.

All subsequent columns correspond to the motion control parameters for that frame. Each of these parameters represents a distinct channel in the MetaHuman Control Rig, and their naming should match the rig control names (e.g., Mouth_LipsPress_L, NoseWrinkle_L, or BrowDown_R). This makes data mapping easier once imported into Unreal Engine.

2.5 Value ranges and data conventions

Each parameter column uses value conventions appropriate to the control it drives:

- **Facial Action Units (AUs):** values range from 0.0 (neutral or inactive) to 1.0 (maximum activation). These values represent normalised magnitudes of muscle activation or blend-shape influence.
- **Head rotation:** defined in degrees across three axes: X, Y, and Z, typically ranging from -180° to $+180^{\circ}$, corresponding to pitch, yaw, and roll. A value of 0 across all axes corresponds to a neutral, forward-facing pose.
- **Eye movement:** defined as floating-point values on the horizontal (X) and vertical (Y) axes, where 0.0 represents a neutral forward gaze, and values range between -1.0 and +1.0, proportionally controlling gaze direction.

Additional columns can include metadata such as AvatarID, FrameRate, or OutputName, which provide context for batch rendering and allow for easier data analysis workflows. These values are not required for animation playback but are useful for managing large datasets and ensuring clear mapping between source data, Unreal Engine sequences, and rendered video outputs.

2.6 Temporal sampling and frame rate

Timing in this framework is frame-based rather than time-based. Each row corresponds to one frame at the target playback framerate. For example, an animation intended to play at 60 frames per second (fps) and last 4 seconds will contain 240 rows per clip. The framerate value itself is applied during Unreal Engine playback (as described in section *Animation Capture Settings*). This approach ensures that motion data remain consistent and engine independent.

2.7 Handling multiple clips in a single file

To improve efficiency, animation data for multiple clips can be stored within the same .csv file. The VideoNumber column differentiates each clip, enabling the present Unreal Engine framework system to segment and iterate through them automatically. This structure simplifies batch processing when generating large stimulus sets, as it consolidates hundreds of clips into

a single file while maintaining discrete animation sequences. See *Table 2.1* for an example .csv structure. In this example, each block of five rows represents a separate animation clip (VideoNumber 0, 1, and 2), each lasting five frames. The AU columns describe the activation levels of four animated parameters across time.

VideoNumber	FrameNumber	AU1	AU2	AU3	AU4
0	0	0.0	0.0	0.0	0.0
0	1	0.02	0.0	0.0	0.0
0	2	0.08	0.0	0.0	0.0
0	3	0.1	0.0	0.0	0.0
0	4	0.2	0.0	0.0	0.0
1	0	0.0	0.0	0.0	0.0
1	1	0.0	0.7	0.0	0.0
1	2	0.0	0.75	0.0	0.0
1	3	0.0	0.8	0.0	0.0
1	4	0.0	0.9	0.0	0.0
2	0	0.0	0.0	0.0	0.0
2	1	0.0	0.0	0.1	0.25
2	2	0.0	0.0	0.25	0.5
2	3	0.0	0.0	0.3	0.75
2	4	0.0	0.0	0.5	1.0

Table 2.1: Example CSV schema illustrating the frame-based animation structure used in the procedural pipeline. Each row represents a single animation frame, VideoNumber segments separate clips within the same file, FrameNumber indexes temporal progression within each clip, and subsequent columns correspond to MetaHuman rig control channels, whose values specify the level of activation for each AU at that frame.

2.8 Recommended practices for data management

When using this format at scale, it is good practice to:

- Maintain consistent naming conventions between .csv headers and Unreal Engine rig inputs to avoid mapping errors.
- Store .csv files with a clearly structured name hierarchy (e.g., by character or experimental condition).
- Include metadata fields such as all animation curve generation parameter values to ensure that all animations can be exactly reproduced.

The .csv schema therefore acts as both a data interface and an experimental record, preserving the precise timing, motion parameters, and randomised elements used to produce each

rendered animation. For full reference on available MetaHuman control rig attributes and naming conventions, see the official Unreal Engine MetaHuman documentation (Epic Games, 2025).

2.9 Importing Data into Unreal Engine

The generated .csv data were imported into Unreal Engine using the DataTable system (a framework for storing structured datasets within the editor). A Blueprint Structure asset (a template defining the variable names and data types expected in the table) was first created, containing variables corresponding exactly to the CSV column names and data types (e.g., VideoNumber [integer], FrameNumber [integer], AU1 [float], etc.). A DataTable asset was then generated from this Blueprint Structure and populated with data from the .csv file through the “Reimport” function. This ensured all animation data were correctly imported and accessible within the engine.

To apply the imported animation data to a MetaHuman character, an Animation Blueprint (a visual scripting framework for defining how animations are applied to a skeletal mesh in real time) was created with the MetaHuman base skeleton. In the Anim Graph (the portion of the Animation Blueprint responsible for blending and driving pose updates) of this blueprint, a Control Rig node (which exposes data input points for the rig’s deformers) was added; all desired rig controls (AUs) were exposed as input pins. Matching float variables were created within the Animation Blueprint and linked to the corresponding Control Rig inputs; integer variables were also created for holding information about the VideoNumber and FrameNumber. Maintaining consistent naming conventions between CSV columns, Control Rig pins, and Blueprint variables was essential for clarity and error prevention. In simple terms, this Blueprint setup allows us to take the imported animation data, now stored in the DataTable asset inside the engine, and use it to animate the MetaHuman AUs through the Control Rig (see *Fig.2.1*).

In the Event Graph, the Animation Blueprint was configured to iterate through the DataTable rows frame by frame (see *Fig.2.2*). A Set Timer by Event node triggered the Fire Anim Frame function at the desired framerate, which updated our created variables’ values on each frame, thereby driving the Control Rig to which they are connected and updating the MetaHuman’s pose in real time (*Fig.2.1*). The VideoNumber and FrameNumber columns controlled recording logic for the Take Recorder, which was initiated at the first frame of each clip and terminated at the final frame. This produced segmented Animation Sequences that could later be rendered as video stimuli.

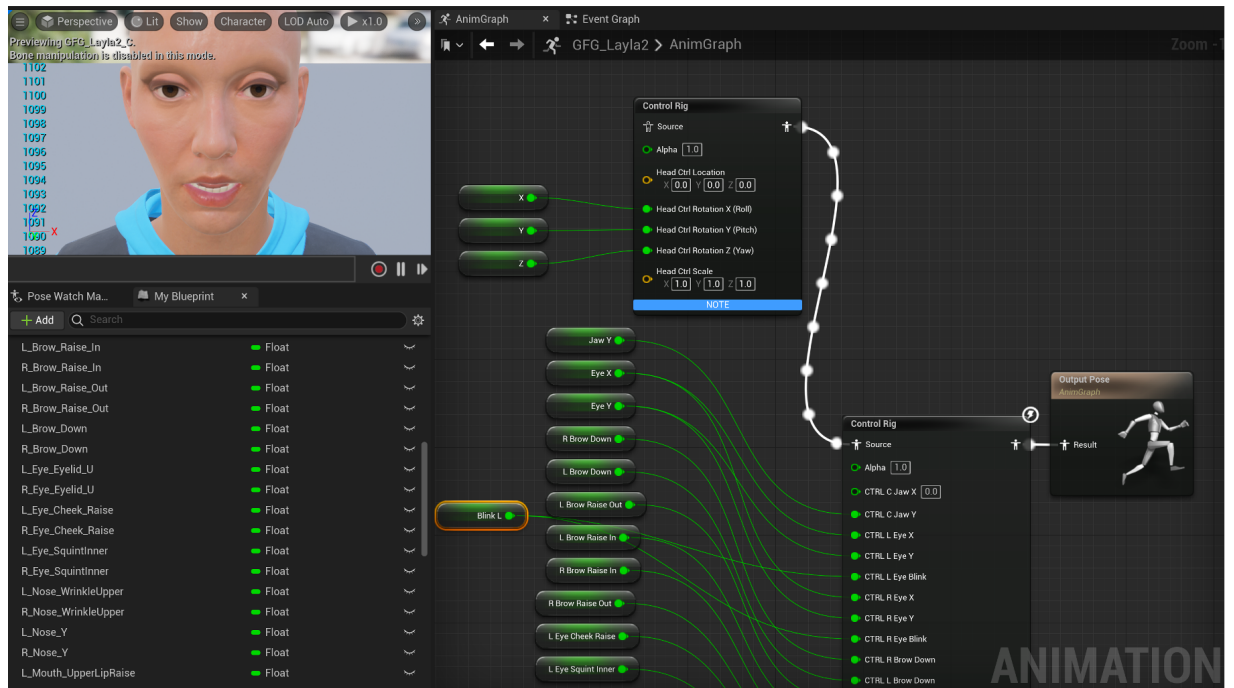


Figure 2.1: Example Animation Blueprint setup for driving MetaHuman facial controls from imported CSV data. The AnimGraph shows Control Rig nodes with exposed input pins corresponding to Action Unit (AU), eye, and head rotation variables. These variables are updated frame-by-frame from the DataTable and routed into the MetaHuman Control Rig, enabling real-time application of externally generated motion parameters to the character's facial mesh.

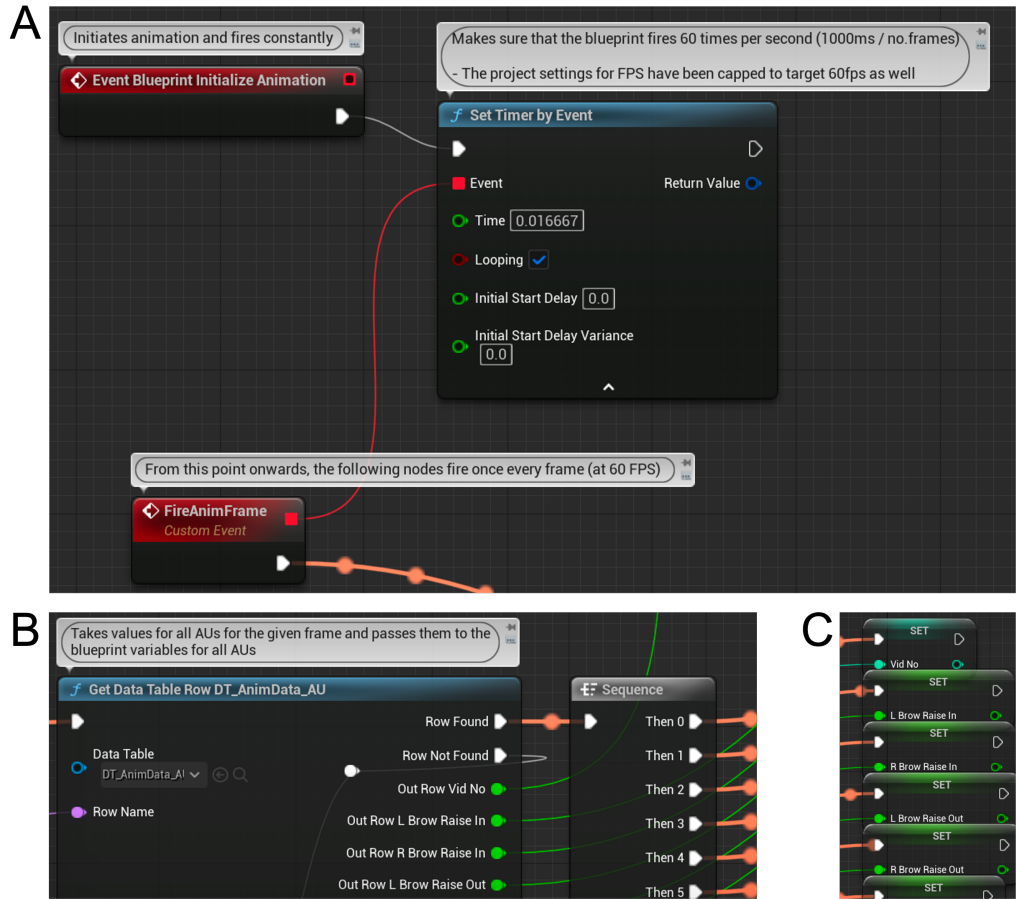


Figure 2.2: Frame-based iteration of animation data within the Animation Blueprint Event Graph. (A) A Set Timer by Event node is configured to fire at 60 frames per second, triggering the FireAnimFrame event at a fixed rate. (B) Each time the event fires, the system reads the corresponding row from the animation DataTable, retrieving the Action Unit (AU) activation values for the current frame. (C) These values are assigned to the Animation Blueprint variables and passed to the MetaHuman Control Rig, updating AU activations and facial pose in real time. Collectively, this setup enables the DataTable to be read sequentially, one row per frame, ensuring frame-accurate playback of externally generated motion data at the specified framerate.

2.10 Animation Capture Settings

To ensure frame-accurate animation playback, the Unreal Engine project settings were configured to operate at a fixed frame rate (e.g., 60 fps) under Engine → General Settings → Use Fixed Frame Rate. Within the Take Recorder, the target MetaHuman character was tracked and recorded at this rate, ensuring a one-to-one correspondence between input frames from the CSV and output frames in the captured Animation Sequences.

2.11 Rendering and Post-Processing

Rendered outputs were generated using Unreal Engine’s Movie Render Queue (MRQ; the high-quality rendering subsystem used for cinematic or research-grade output), which allows for high-quality batch rendering. By default, MRQ produces image sequences (e.g., PNG or EXR), consistent with workflows used in film and visual effects production. These sequences can be compiled into video files externally using tools such as FFmpeg or video editing software. FFmpeg is an open-source command-line multimedia framework widely used for processing, encoding, decoding, and converting audio and video files across different formats (Tomar, 2006). Alternatively, Unreal Engine can be configured to interface directly with FFmpeg to export video files automatically without additional steps outside of the engine.

While Unreal Engine’s FFmpeg integration requires additional configuration, it provides an efficient and streamlined solution for producing video outputs suitable for experimental stimuli generation at scale.

2.12 Further Automation

For advanced users, Unreal Engine’s Python Editor Scripting API can be used to automate this process further, from data import to rendering. In principle, this approach can bypass the Take Recorder and Animation Blueprint stage, generating animation sequences directly from the .csv animation data. However, the implementation described in this thesis prioritised accessibility through Unreal Engine’s base tools, demonstrating the feasibility of this research workflow as a proof of concept. Researchers are encouraged to expand and iterate on this approach, for example, by streamlining the workflow with Python Editor Scripting. The presented system logic and data pipeline serves as a foundation for these future developments.

2.13 Technical Considerations and Limitations

The described pipeline requires substantial computational resources. Running and rendering high-fidelity MetaHumans in Unreal Engine demands a modern GPU and sufficient memory to maintain stable performance. For efficient playback and animation capture, scalability settings can be temporarily reduced to ‘Low’, whereas rendering through the Movie Render Queue should use the highest (‘Cinematic’) settings to achieve maximum visual fidelity. Rendering time is largely dependent on GPU performance; faster GPUs result in shorter render durations.

The workflow was developed and tested on a Windows 11 desktop equipped with an NVIDIA

RTX 3070 GPU (8GB VRAM), Ryzen 5 5600 CPU, and 32GB RAM. Tests on macOS systems revealed several limitations at the time of development: Unreal Engine’s Groom Hair system was not fully supported, preventing rendering of hair, eyelashes, and fine facial detail at high settings. Additionally, direct FFmpeg integration inside Unreal Engine for video export was not available on macOS at the time of writing, requiring manual post-processing in video editing software or additional scripting outside of Unreal Engine for video rendering automation.

The pipeline also demands significant storage capacity. MetaHuman assets and Unreal Engine project files occupy large amounts of disk space (up to 3GB per MetaHuman character asset) and rendered sequences or video outputs can accumulate quickly depending on resolution and codec/format choices. Installing the engine itself can amount to 100GB or more. Researchers should evaluate whether such graphical fidelity is necessary for their experimental goals and ensure sufficient storage and backup capacity.

Despite some of the technical caveats, once configured, the pipeline offers excellent scalability. It enables automated generation of thousands of animations with minimal manual intervention, making it ideal for large-scale perceptual experiments such as reverse correlation studies. Although setup and rendering require substantial resources, the system provides a robust, high-fidelity, and reproducible foundation for behavioural research with digital humans.

2.14 Chapter Summary

This chapter outlined the development and implementation of a custom animation pipeline for generating realistic, data-driven MetaHuman stimuli within Unreal Engine. The framework bridges Python-based motion parameterisation with Unreal Engine’s Control Rig and Sequencer systems, allowing researchers to precisely control facial, head, eye, and body motion while maintaining the photorealistic rendering quality of modern digital humans. Once established, this workflow enables large scale, reproducible generation of dynamic stimuli suitable for experimental applications in psychology, neuroscience, and human-computer interaction.

Beyond serving the immediate requirements of this thesis, the pipeline demonstrates the broader feasibility of using state-of-the-art real-time graphics engines for empirical research. The modular structure of the pipeline, comprising distinct components for data generation, import, rig control, and rendering, makes it easily adaptable to other research paradigms, from reverse correlation and emotion perception to social cognition. While computationally demanding, the framework balances visual fidelity with experimental precision, establishing a robust methodology for future digital human studies.

The following chapter applies this pipeline to its first experimental context: validating

whether MetaHuman avatars can serve as scientifically credible stimuli for psychological research. Study 1 compares MetaHumans with the established Generative Face Grammar (GFG; Yu et al., 2012) system across perceptual dimensions such as humanness, realism, eeriness, and familiarity, thereby assessing both the perceptual realism and affective validity of the MetaHuman framework as a foundation for subsequent studies of the thesis on perception of natural neutral motion in digital human characters.

Chapter 3

Validating MetaHuman Avatars: A Mixed Methods Benchmark (Study 1a & 1b)

3.0.1 Chapter Overview

This chapter presents the first empirical component of the thesis, evaluating the suitability of Epic Games’ MetaHuman framework (Epic Games, 2021, 2025) as a research tool for studying face perception. While industries such as film, games, and virtual production now routinely employ photorealistic digital humans (Xu & Zhang, 2025), social face perception research has largely continued to rely on highly influential but older stimulus systems such as the Generative Face Grammar (GFG; Jack & Schyns, 2017; Yu et al., 2012) and FaceGen (Blanz & Vetter, 1999; Oosterhof & Todorov, 2008). Establishing the scientific validity of modern, high-fidelity platforms is therefore an essential step toward improving the ecological validity and interpretive relevance of experimental findings towards real human face perception. Specifically, we test whether modern, high fidelity MetaHuman characters elevate humanness, realism, and familiarity ratings without a corresponding increase in eeriness compared to GFG characters.

The chapter comprises of two complementary studies. *Study 1a* provides a quantitative comparison between the GFG and MetaHuman avatars using established perceptual measures of humanness, eeriness, realism, and familiarity derived from the uncanny valley literature (Ho & MacDorman, 2010, 2017). *Study 1b* expands on this through a qualitative thematic analysis of participants’ subjective reflections, investigating how observers describe the visual and behavioural features that make digital characters appear more or less “lifelike”. Together, these studies offer converging statistical and experiential evidence for the perceptual credibility of MetaHumans as research stimuli, setting the foundation for the motion-based experiments in subsequent chapters.

3.1 Study 1a: Quantitative Benchmark of MetaHuman and GFG Avatars

3.1.1 Introduction

As established in the general introduction (*Chapter 1*), advances in rendering technology and digital human animation have transformed the realism of avatars used throughout the entertainment industry, while psychological research has been slower to adopt equivalent stimulus platforms. The present chapter therefore evaluates whether MetaHuman avatars provide perceptual advantages over the established GFG framework before they are adopted throughout the remainder of this thesis.

3.1.1.1 Limitations of Experimental Stimuli

While the entertainment industry has embraced cutting-edge rendering and animation pipelines, experimental psychology has often relied on three-dimensional (3D) avatar systems whose visual rendering and graphical fidelity lag behind contemporary digital human technologies, even though the underlying facial modelling and animation control remain scientifically powerful. In this chapter, when referring to stimuli, we specifically refer not to static photographs or video recordings of real faces, but to controllable 3D avatars whose geometry, texture, and motion can be systematically manipulated. Such avatars are essential for paradigms that require precise parametric control, enabling researchers to isolate the contribution of individual features, expressions, or movements while holding other aspects constant (Jack & Schyns, 2017).

The most influential 3D facial stimulus frameworks used in psychological research to date include FaceGen (Blaiz & Vetter, 1999; Oosterhof & Todorov, 2008) and the Generative Face Grammar (GFG; Yu et al., 2012). These tools were groundbreaking in that they allowed systematic manipulation of facial appearance and dynamics through underlying morphable face models and facial action unit (AU)-based animation, where the Facial Action Coding System (FACS) decomposes facial movement into combinations of AUs corresponding to specific muscle activations (Ekman & Friesen, 1978). Both FaceGen and the GFG derive their stimulus spaces from scans of real human faces: FaceGen is based on a statistical 3D morphable model constructed from laser-scanned heads, whereas the GFG relies on FACS-based AU blendshapes and an identity space derived from 4D stereo scans of FACS-trained actors together with additional scanned identities, enabling the synthesis of novel facial identities and expressions. Although both frameworks have been used to investigate how people perceive social traits in relation to facial morphology and texture (Gosetti et al., 2023, 2024; Hensel et al., 2020; Todorov, 2008;

Todorov et al., 2008, 2011, 2025; Zhan et al., 2020, 2021), the GFG has also been applied in landmark reverse-correlation studies of emotional expression (Jack, Garrod, et al., 2012; Jack et al., 2014; Yu et al., 2012) and influential work on other dynamic facial movements (Chen & Jack, 2017; Chen et al., 2015, 2018; Gill et al., 2014; Richoz et al., 2015). Thanks to the precise experimental control these stimulus frameworks provide, they remain central to work on emotion perception and dynamic facial social signalling (Jack & Schyns, 2017).

However, despite their scientific utility, both FaceGen and the GFG are products of an earlier generation of 3D rendering technology, meaning that they produce stimuli that lack the graphical fidelity of contemporary digital humans used across the media and entertainment industries. Neither FaceGen or GFG stimuli implement modern 3D graphics solutions such as advanced PBR-based advanced materials, or subsurface scattering, resulting in flat, opaque-looking skin (McDonnell et al., 2012). Furthermore, the texture resolution and detail, and topology polygon counts are significantly lacking by today's standards, resulting in jagged and sharp edges in parts of the 3D models; this is particularly noticeable with the eyelid, ear, and lip topology across the two systems. Both systems also implement overly simplified approaches towards hair rendering. For example, neither system renders the characters with any eyelashes, in both cases the eyebrows are simply included in the face texture itself leading to a flat, non-realistic look. The head hair rendering is also significantly limited, in both cases stimuli tend to be all bald characters. Although the GFG is also capable of rendering hair with a single short, buzz-cut-like hairstyle in various colours; however, this is also far behind modern hair rendering solutions. Thus, although these digital human avatars provide valuable experimental control, their overall appearance and motion can appear synthetic or stylised compared to modern state-of-the-art solutions in digital human characters. These technical constraints raise concerns not only about the aesthetic quality, but also about the ecological validity of perception research that aims to generalise effects found using 3D avatars onto perceptions of real human beings.

3.1.1.2 Ecological Validity and Realism

The fundamental goal of research on perception of social signals is to understand how people interpret and respond to various social cues conveyed by other humans in real-world situations. For findings of such research to generalise beyond the laboratory, the stimuli used must elicit the same, or at least comparable, cognitive and behavioural processes that underlie real social interactions. This requirement is reflected in the concept of ecological validity, which refers to the extent to which the contextual, sensory, and behavioural cues of an experimental setup reflect those present in natural human interaction (e.g., Blascovich et al., 2002; Kingstone et al., 2008).

However, as Holleman et al. (2020) argue, ecological validity is often raised in an overly

broad or vague manner, typically used as a criticism of laboratory paradigms without clearly specifying which aspects of “real life” are in question. In line with their call for greater specificity, we therefore use the term in a more restricted and context-specific sense. More specifically, the context here is the perception of digital humans: we take an ecologically valid digital avatar to be one whose appearance and motion most closely resemble real humans to engage comparable perceptual and cognitive mechanisms (e.g., in terms of facial morphology, facial deformation, rendering, and lighting). Behavioural work with FaceGen avatars indicates that they are not psychologically equivalent to real faces. In a series of face recognition experiments, Crookes et al. (2015) found that classic markers of face expertise, such as the own-race advantage (better recognition memory for faces from one’s own racial group) and the other-race effect (poorer discrimination for other-race faces), were markedly reduced or even absent for FaceGen stimuli relative to photographs of real people. Crookes and colleagues specifically argue that these computer-generated faces still look artificial; lacking the fine-grained surface texture, imperfections, and cues to animacy characteristic of real humans. They hypothesise that this likely explains why the FaceGen stimuli do not fully recruit the typical expert face-processing mechanisms engaged by real human faces.

As discussed in Chapter 1, psychophysical and graphics research demonstrates that render style and animation quality strongly influence how viewers perceive virtual human characters (Kätsyri et al., 2015; McDonnell et al., 2012; Wisessing et al., 2020). Furthermore, Higgins et al. (2021) compared high and low-fidelity versions of MetaHuman characters that were matched in identity, motion, voice, lighting, and rendering, and found that high-fidelity avatars were seen as more humanlike, less eerie, and more attractive. Critically, the high fidelity MetaHumans were perceived as expressing the same emotions with greater intensity than their low fidelity counterparts. Taken together, these findings suggest that the fidelity and coherence of a character’s appearance, particularly surface rendering and lighting, are not just aesthetic details but core determinants of how observers perceive lifelikeness, emotion, and by extension, the ecological validity of a given avatar, especially when studying emotional expression. Methodologically, this implies that studies probing perception of emotion or social traits with low fidelity or poorly rendered characters risk confounding their measures with artefacts of character fidelity, thus, limiting the extent to which results can be generalised to real human faces.

In sum, the ecological validity of psychological research utilising digital humans relies on adopting avatars whose appearance and behavioural fidelity are generally on par with the state-of-the-art realism of digital human characters that people encounter across contemporary media nowadays. As observers are increasingly exposed to highly realistic facial animation and rendering across digital platforms, psychology must evolve its tools accordingly if it is to build models of social perception grounded in lifelike behaviour rather than confounded responses to outdated graphics technologies. These issues become especially salient when considering theories of the

uncanny valley, where mismatches between appearance and motion fidelity are thought to play a central role.

3.1.1.3 The Uncanny Valley and Mismatch Theory

The uncanny valley refers to a phenomenon first proposed by Masahiro Mori (1970; Mori et al., 2012), in which increasing human likeness of artificial agents generally leads to more positive reactions towards them, until a point at which near-human but imperfect realism instead evokes discomfort or eeriness. This point, where responses dip, is the metaphorical ‘valley’ in the relationship between human likeness and affective response. Essentially, as appearance becomes almost, but not quite, human, small deviations can trigger a sharp drop in comfort. Mori also hypothesised that movement would exacerbate this effect: once an almost-human agent residing in the uncanny valley begins to move, it can elicit even stronger negative affect toward the agent. The curve is typically assumed to rise again once realism is sufficiently high and perceptual cues are fully consistent, such that extremely humanlike artificial faces can theoretically approach the reactions elicited by real human faces without uncanny feelings arising.

Contemporary theories have refined Mori’s idea by identifying the specific perceptual conflicts that produce the uncanny response. Rather than realism alone, it is now thought that the uncanny valley arises when there is a mismatch between perceptual cues that should typically align in real human perception (Kätsyri et al., 2015). For example, an avatar may display anatomically realistic skin and eyes but move in a robotic and unnatural manner. Alternatively, motion may appear fluid and lifelike, but the graphical rendering may fail to capture the physical properties of human skin. Most empirical work so far has focused on mismatches between different static cues (e.g., realistic textures combined with unrealistic proportions), and Kätsyri et al. (2015) note that, as far as they are aware, the specific proposal that negative affinity is driven by inconsistencies between static appearance and dynamic motion cues has not yet been systematically examined. More generally, the mismatch account suggests that eeriness results not from realism as such, but from inconsistencies among realism cues.

Several empirical studies support this mismatch account. Seyama and Nagayama (2007) showed that artificial faces were judged most eerie not when they were simply low in realism, but when realistic surface textures were combined with atypical or non-human facial proportions, suggesting that conflicting appearance cues drive negative evaluations. Similarly, MacDorman et al. (2009) manipulated features such as eye size and skin texture in computer-generated faces and found peaks in eeriness when highly realistic textures were paired with subtly abnormal feature configurations, again pointing to inconsistency among static visual cues rather than realism per se. Extending this work beyond vision alone, Mitchell et al. (2011) demonstrated that mismatching the human realism of face and voice (e.g., a robot face with a human voice, or a

human face with a synthetic voice) reliably increased eeriness compared with matched conditions, indicating that cross-modal inconsistencies in human realism can also produce an uncanny valley. More recently, Chattopadhyay and MacDorman (Chattopadhyay & MacDorman, 2016; MacDorman & Chattopadhyay, 2016) used categorisation and rating tasks on precisely aligned continua between 3D models and photographs, selectively mismatching the realism of different feature sets (e.g., eyes and mouth vs. skin), and showed that the eeriest stimuli were those with clearly artificial yet internally inconsistent human features, with such effects emerging for humans and animals but not for objects.

Taken together, these findings imply that perceptual realism depends not only on detailed geometry and textures, but on a coherent match between appearance, motion, and behavioural expressivity. This is especially clear in work showing that static appearance realism can shape how dynamic cues are read: when Higgins et al. (2021) presented MetaHuman faces animated with identical captured facial performances at different levels of visual fidelity, and found that higher-fidelity characters were judged not only more humanlike, attractive and less eerie, but also as expressing more intense happiness and sadness than their lower-fidelity counterparts. Although this study did not explicitly manipulate mismatches between character realism (low vs. high) and animation realism (low vs. high), the finding that identical motion is perceived differently as a function of visual fidelity is consistent with mismatch accounts; making the face more humanlike appears to bring static and dynamic cues into closer alignment, thereby enhancing emotional intensity and realism, and reducing eeriness. This has direct implications for studying motion dynamics using avatars with different realism levels, where the interplay between visual fidelity and animation quality may determine whether observers perceive a character as natural, mechanical, or uncanny, and whether the findings are generalisable to real human expressions and behaviour.

Despite the growing body of work on the uncanny valley, several important questions remain unresolved. Much of the empirical literature has focused primarily on manipulations of static appearance, such as facial proportions, texture realism, or feature configuration, while comparatively little research has examined how subtle dynamic behaviours contribute to uncanny impressions. In particular, the role of naturalistic facial motion and idle behaviour remains underexplored, even though Mori (1970) originally suggested that movement may intensify uncanny responses. Furthermore, many existing studies rely on avatar systems whose graphical fidelity differs substantially from contemporary digital human characters, making it difficult to disentangle perceptual effects driven by behavioural realism from those arising from limitations in rendering quality. Addressing these issues requires experimental frameworks that combine high-fidelity digital humans with precise parametric control over facial motion, allowing researchers to systematically manipulate behavioural cues while controlling for appearance effects.

3.1.1.4 Measuring the Perception of Digital Humans

To evaluate how observers respond to digital human avatars, it is necessary to use psychometric instruments that distinguish how they look from how they make people feel. Established measures such as the Godspeed questionnaire Bartneck et al. (2009) developed indices for constructs such as anthropomorphism and likeability. However, later work showed that these scales tend to move together rather than capture distinct dimensions (Ho & MacDorman, 2010). In practice, higher “anthropomorphism” scores usually coincide with higher “likeability” and other aspects of general positive evaluation, making it difficult to tell whether participants are responding to human-like appearance specifically or simply expressing that they like the agent. As a result, these measures cannot cleanly separate perceived human-likeness from feelings of comfort or discomfort and are not well suited for testing hypotheses derived from Mori’s (1970) uncanny valley framework, which assumes at least partly independent dimensions of humanness and eeriness. A recent meta-analysis of uncanny valley research supports this notion (Diel et al., 2021). It shows that generally positive ratings (such as likeable or pleasant) are relatively insensitive to the characteristic “valley” shape of the relationship, especially for less humanlike agents, and are easily influenced by factors such as attractiveness, context, or narrative framing. In other words, general likeability ratings tend to reflect overall appeal rather than the specific uncanny discomfort that Mori’s framework aims to capture.

Ho and MacDorman (2010) therefore developed an alternative set of indices that aim to measure perceived humanness, eeriness, warmth, and attractiveness as distinct dimensions; that is, scales that do not simply rise and fall together, but show relatively low overlap and capture different psychological responses. This approach improves on the Godspeed questionnaire, where anthropomorphism and likeability tend to load on the same underlying factor (Ho & MacDorman, 2010). The humanness index captures the extent to which a character is seen as human rather than artificial (e.g., artificial-natural, synthetic-real, inanimate-living, mechanical movement-biological movement). This scale was designed to track perceived human realism, across both robots and computer-animated faces, without collapsing into a general “likeability” factor. In later refinements of their 2010 indices, Ho and MacDorman (2017) showed that a revised set of humanness items maintains high internal reliability while better capturing graded differences in perceived human likeness across the full spectrum, rather than clustering stimuli simply into “robot-like” and “human-like” groups.

On the other hand, the eeriness index targets the affective element central to uncanny valley accounts: the unusual discomfort or unease that can arise when an entity looks or moves almost, but not exactly, like a human. Ho and MacDorman (2010, 2017) conceptualise eeriness as multi-faceted, with items loading on many different subdimensions (e.g., reassuring-eerie, dull-freaky, plain-weird, bland-uncanny). Most importantly, the final eeriness index is statistically

decorrelated from humanness and from general warmth/likeability (Ho & MacDorman, 2010, 2017), allowing detection of agents that are simultaneously very humanlike and very eerie unlike the Godspeed questionnaire. This is the exact pattern predicted at the bottom of the metaphorical uncanny valley and by mismatch theories (Kättsyri et al., 2015).

The meta-analysis by Diel et al. (2021) provides converging evidence that the measures described above are most suitable for examining the uncanny valley phenomenon. Across 72 studies, they found that uncanny valley-specific negative affect terms and indices (e.g., eerie, uncanny, creepy, threatening, to name a few) produced some of the largest and most reliable effect sizes, particularly on the nonhuman side of the valley; that is, agents that are still clearly artificial but increasingly humanlike. Whereas, positive warmth/likeability measures were much less effective at capturing the valley-shaped dip in the relationship between affect and perception. On this basis, Diel and colleagues recommend using specific negative uncanny items as primary outcome measures and treating general warmth/likeability as secondary, more confounded dimensions. The Ho and MacDorman (2010, 2017) eeriness indices are very closely in line with these design principles and are repeatedly highlighted (directly or in derived form) in the studies included in the meta-analysis.

A key contribution of the 2017 refinement is that it explicitly addresses categorical perception in how people judge robots and animated characters. Ho and MacDorman (2017) found that the earlier (2010) versions of the scales tended to collapse stimuli into two tight clusters (robots vs computer-animated humans), despite substantial variation within each group. By using card sorting and laddering interviews to derive new adjective pairs, and then validating them in a large web survey, the revised (2017) indices produce a broader spread of ratings across humanness and eeriness, including in the ambiguous boundary region between “robot” and “human”. When humanness is plotted against (reverse scaled) eeriness, the resulting curve more closely resembles Mori’s proposed uncanny valley, with a clear dip in affective comfort for entities that are almost (but not fully) human. This makes these indices particularly appropriate for examining how visual fidelity and/or motion realism contribute to uncanny responses, and their effectiveness is consistent with the patterns reported by Diel et al., (2021).

In addition to humanness and eeriness, many uncanny valley studies include realism and familiarity as complementary rating dimensions (e.g., Chattopadhyay & MacDorman, 2016). Realism ratings are typically intended to index the perceived believability or photographic quality of the stimulus, largely independent of whether that realism is experienced positively or negatively. Familiarity ratings tap into how common, recognisable, or weird the stimulus feels, and thus help distinguish agents that are perceptually convincing but affectively unsettling (high realism, high humanness, high eeriness, low familiarity) from those that are both realistic and comfortable in a social sense. Diel et al. (2021) show that such constructs (e.g., familiarity, aesthetics, animacy/experience) can produce considerable effects when operationalised with well

targeted items, but also emphasise that they are less specific to the uncanny valley than the more specific negative affect terms such as eerie, creepy, or uncanny. Accordingly, familiarity is best interpreted alongside more uncanny valley-specific scales in a complementary manner, rather than used as the sole outcome measure.

Taken together, these measures are well suited to the present study, which compares GFG-based and MetaHuman-based avatars along core uncanny valley dimensions. By adopting humanness and eeriness indices developed and refined by Ho and MacDorman (2010, 2017), in line with the methodological recommendations from the Diel et al. (2021) meta-analysis, alongside complementary realism and familiarity ratings, we can test whether higher fidelity MetaHuman characters are perceived as more humanlike and realistic without a corresponding increase in eeriness. This provides a solid framework for evaluating whether modern digital humans such as MetaHumans can move observers closer to an experience comparable to real human perception without falling into the depths of the uncanny valley when compared to the older digital human frameworks such as the GFG.

3.1.1.5 Study Rationale and Aims

Taken together, the previous sections highlight a growing gap between the technological realism available in contemporary digital media and the stimulus fidelity typically used in psychological research. While entertainment industries nowadays increasingly adopt photorealistic digital humans rendered with state-of-the-art animation and rendering techniques, many experimental paradigms continue to rely on stimuli produced using more dated approaches such as FaceGen or the Generative Face Grammar (GFG). Although these tools were instrumental in advancing research on facial perception and emotion dynamics (Dawel et al., 2022; Dobs et al., 2018; Jack, Garrod, et al., 2012; Jack & Schyns, 2017; Oosterhof & Todorov, 2008; Todorov et al., 2011; Yu et al., 2012), their graphical fidelity, particularly in surface rendering and lighting, no longer reflects the perceptual standards set by modern graphical solutions. This technological lag raises important questions about the ecological validity and generalisability of findings based on these platforms: if participants routinely engage with highly realistic digital humans in games, films, and virtual media, their perceptual expectations are likely more aligned to that standard. Consequently, results derived from lower-fidelity stimuli may underestimate or distort responses to truly lifelike faces.

The present study directly addresses this issue by evaluating Epic Games' MetaHuman framework as a tool for psychological research on facial perception. MetaHumans combine state-of-the-art real-time rendering techniques with industry-grade facial animation rigging that supports fine-grained control of expressions and motion. At the same time, the system allows for the intricate parametric controllability that made platforms such as the GFG valuable for

experimental work, enabling researchers to manipulate facial expression, structure, and motion in a systematic and reproducible way.

The central aim of this study is to empirically compare the GFG and the MetaHuman avatars across key perceptual dimensions (humanness, eeriness, familiarity, and realism) measured using validated scales from Ho and MacDorman (2010, 2017) and based on recommendations by Diel et al. (2021). Specifically, we test whether modern, high fidelity characters elevate humanness, realism, and familiarity ratings without a corresponding increase in eeriness. Thus, avoiding the perceptual pitfalls predicted by the uncanny valley hypothesis (Mori, 1970; Kätsyri et al., 2015). Demonstrating that MetaHumans can deliver greater ecological validity without compromising affective comfort would help bridge the gap between experimental control and perceptual believability, and provide a foundation for future research on emotion, face perception, and social cognition using state-of-the-art digital humans. Based on this, we hypothesise that MetaHumans will be seen as more humanlike, more realistic, and more familiar, whilst being perceived as less eerie than the GFG avatars.

3.1.2 Methods

3.1.2.1 Participants

Participants were recruited through leaflet advertisements around the University of Glasgow and word-of-mouth opportunity sampling. The participation criteria included being aged 18 or older, having normal or corrected-to-normal vision, being free of any: (1) lexical, reading, language, hearing atypicality (e.g., dyslexia), (2) social/emotion processing related atypicality (e.g., autism spectrum disorder, depression, anxiety), (3) any face processing atypicality (e.g., prosopagnosia). In total, 47 participants (23 M, 24 F) were recruited with an average age of 30.23 (SD 10.09). All participants met the participation criteria. As compensation, participants were entered into an anonymised prize draw for one of two £50 Amazon vouchers. The study received ethical approval from the University of Glasgow's College of Medical, Veterinary & Life Sciences Ethics Committee (Ethics Code: 200240259).

In addition, a subset of these participants ($n = 44$) continued to the optional second section of the questionnaire, which included a small number of supplementary items and open-ended questions reported in *Study 1b*.

3.1.2.2 Measures

Perceptual responses to each avatar were assessed using validated semantic differential scales widely employed in uncanny valley research (Ho & MacDorman, 2010, 2017; Diel et al., 2021). Each item was rated on a 7-point bipolar scale anchored by opposing adjectives. To minimise participant fatigue given the number of stimuli, a concise subset of items was selected that captured the core dimensions of the original indices provided by Ho and MacDorman:

- **Humanlikeness** was measured with three items (Inanimate-Living, Synthetic-Real, and Human-made-Humanlike).
- **Eeriness** was measured with three items (Dull-Freaky, Predictable-Eerie, and Plain-Weird).

In addition, two supplementary single-item scales assessed:

- **Familiarity** (Strange-Familiar)
- **Realism** (Not realistic-Very realistic)

reflecting overall perceptual believability and social comfort.

For each participant, mean scores were computed per dimension (humanness, eeriness, familiarity, and realism). Higher scores indicated greater perceived humanness, eeriness, familiarity, or realism, respectively.

Additionally, from the optional section of the questionnaire, one item assessing identity similarity between each GFG-MetaHuman avatar pair (*“To what extent do the two avatars look like they could represent the same person?”*) was analysed in the present study as it directly relates to the perceptual comparability of the two avatar frameworks. For this measure, 1 = *“Not at all”* and 7 = *“Definitely”*.

3.1.2.3 Stimuli

The stimulus set comprised of six pairs of avatars, with each pair depicting the same character generated with two different systems: the GFG and MetaHuman. The GFG avatars were drawn from the established library used in previous studies of dynamic facial expression perception (Yu et al., 2012; Jack & Schyns, 2017). The GFG model parameterises facial structure and motion based on a large dataset of 3D facial scans of real humans, enabling the synthesis of new but statistically plausible facial identities derived from that dataset.

To create the MetaHuman counterparts of the GFG characters, the 3D facial meshes from six generated GFG characters were first extracted and imported into Unreal Engine 5.6 Games (2025). The Mesh to MetaHuman tool was then used to reconstruct and transfer each GFG facial topology onto a MetaHuman base model.

Following reconstruction, the MetaHuman Creator interface was used to manually refine surface-level features such as skin tone, texture, eye colour, and hair style, to achieve the closest possible visual match to the original GFG characters. Because the GFG avatars are limited to short, buzz-cut-like hairstyles, the most closely corresponding hair type was applied to all MetaHuman counterparts to ensure stylistic consistency across systems.

The character set was designed to represent three male and three female identities, with each male-female pair corresponding to one of three broad ethnic groups (Caucasian, East Asian, and African; see *Fig.3.1*). This design controlled for potential effects of gender and ethnicity on perceptual judgements of humanness, realism, familiarity, and eeriness. Care was taken to ensure that the paired avatars were as closely matched as possible in identity, gender, ethnicity, and overall facial configuration, such that any observed differences in participant ratings could be attributed primarily to rendering quality and perceptual realism, rather than morphological variation.



Figure 3.1: Stimulus characters used in the experiment. Six identities were first generated using the Generative Face Grammar (GFG) framework and then transferred to the MetaHuman system using the Mesh to MetaHuman tool in Unreal Engine. (A) shows the original GFG avatars, while (B) shows the corresponding MetaHuman counterparts reconstructed from the GFG meshes. Each row represents a matched identity pair, with three male and three female characters spanning three broad ethnic groups (Caucasian, East Asian, and African). The MetaHuman characters were subsequently refined in MetaHuman Creator by manually selecting the closest matching hair style and adjusting facial and eye textures to approximate the appearance of the original GFG identities as closely as possible. The original GFG characters comprised the head and neck only, whereas the corresponding MetaHuman characters included part of the upper torso as MetaHumans come with a full body rig. The characters were framed so that the heads were equally visible across both avatar frameworks.

This process mostly preserved the overall facial geometry of the original GFG characters while adapting them to the MetaHuman rig structure, enabling the characters to benefit from MetaHumans’ higher fidelity deformation, shading, and rendering systems. However, several technical differences between the two frameworks should be noted. While the global facial

shape of the GFG identities could largely be maintained (see *Fig.3.2*), the MetaHuman system uses a substantially denser and more detailed mesh topology than the original GFG models. As a result, certain facial regions are reconstructed with an introduction of additional geometric detail during the Mesh to MetaHuman conversion process.

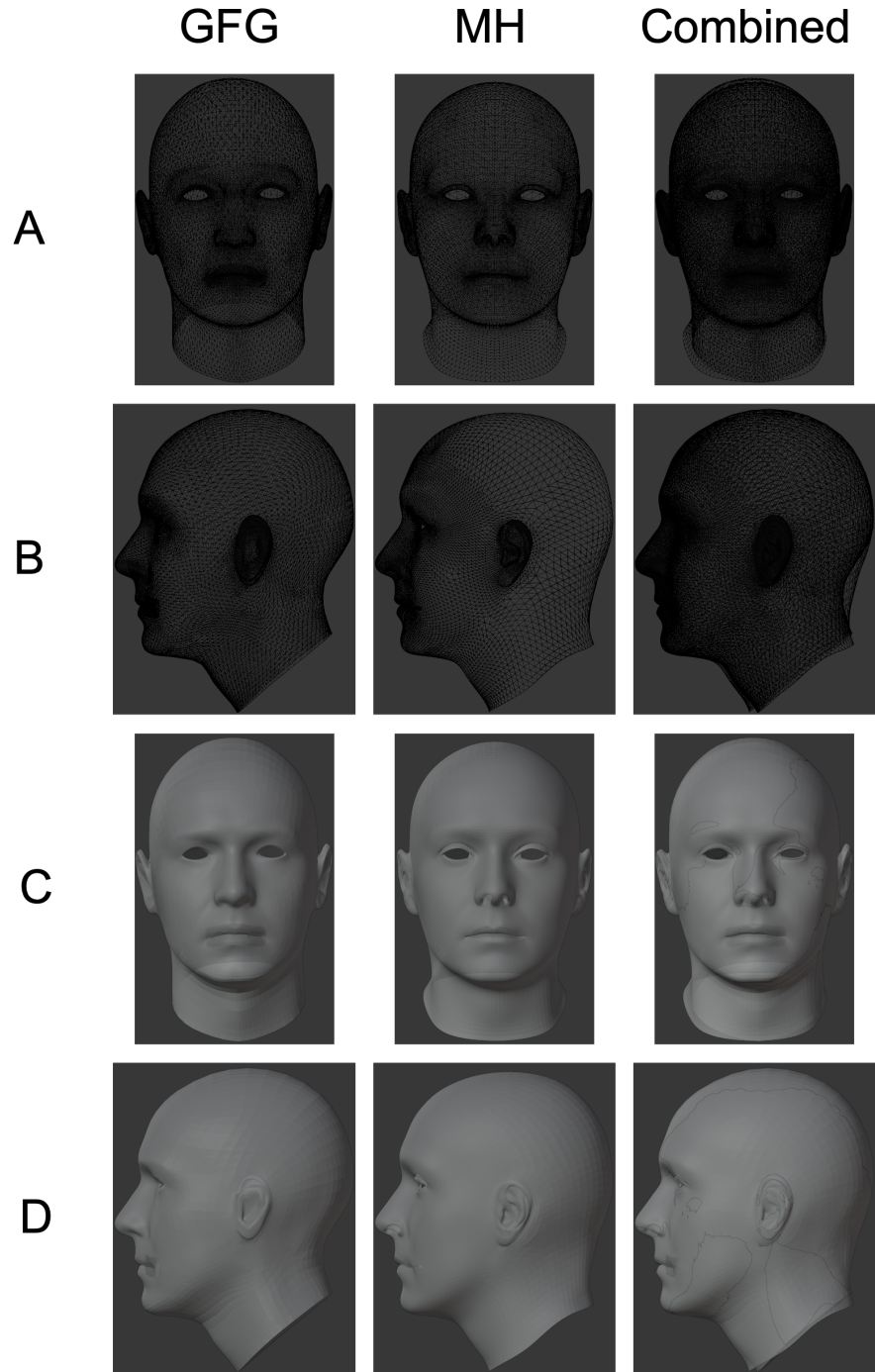


Figure 3.2: Comparison of mesh geometry between the original Generative Face Grammar (GFG) character and its reconstructed MetaHuman (MH) counterpart for the Caucasian male identity. Columns show the GFG mesh, the reconstructed MetaHuman mesh, and an overlay of both meshes for visual comparison. Rows A and B show wireframe views highlighting mesh topology from frontal and lateral perspectives, while rows C and D show the corresponding clay renders illustrating overall surface geometry. The comparison demonstrates that the global facial structure of the identity is largely preserved during the Mesh to MetaHuman transfer process.

For example, the eye region of the GFG meshes contains relatively simple geometry, whereas

MetaHuman characters include more complex anatomical structures such as tear ducts, layered eyelids, and detailed eye-line geometry. When the GFG meshes are converted into the MetaHuman format, these additional structures are introduced by the MetaHuman base topology, resulting in geometric differences in these areas. Similar effects occur in other regions such as the nostrils, where MetaHuman topology represents internal nasal structure with more detail than the GFG meshes. Consequently, while overall identity cues were preserved, some localised facial differences between the two systems are unavoidable in this transition process due to the base topology of the GFG meshes (Fig.3.3).

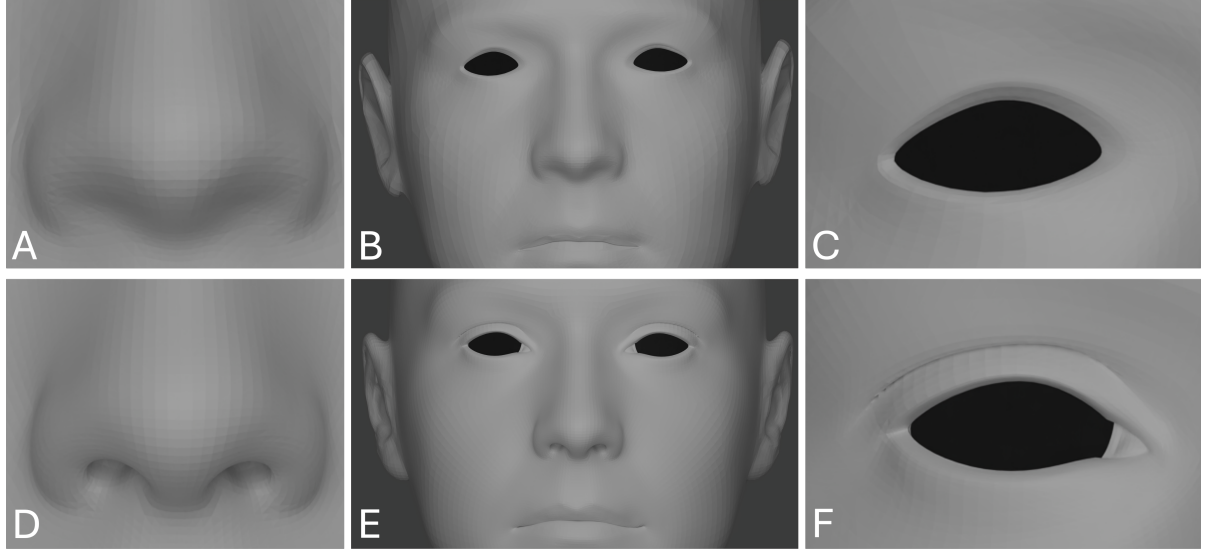


Figure 3.3: Localised topology differences between the GFG and MetaHuman meshes. The top row (A-C) shows the original Caucasian male GFG mesh, while the bottom row (D-F) shows the corresponding Caucasian male MetaHuman reconstruction generated using the Mesh to MetaHuman pipeline. Panels (A) and (D) show the nasal region, (B) and (E) show the close-up of the frontal facial mesh, and (C) and (F) show a close-up of the eye region. Compared to the GFG mesh, the MetaHuman mesh introduces additional anatomical detail, including more defined nostril geometry, layered eyelid structure, and tear duct regions around the eye. These differences arise from the higher resolution topology used by the MetaHuman framework and illustrate how local geometric detail is added during the identity transfer process while preserving the general overall facial structure.

A formal vertex-level comparison of mesh deviations between the two frameworks was not conducted. Such an analysis is complicated by substantial differences in vertex counts and mesh organisation between the models, as MetaHuman assets contain multiple integrated components (e.g., eyes, teeth, eyelashes, and other facial elements) with complex vertex groupings. Isolating all the head meshes and systematically removing these components for analysis would be a substantial technical challenge. Establishing a reliable vertex-to-vertex correspondence between the two mesh structures for quantitative comparison was therefore beyond the scope of the present study.

Additionally, texture replication was constrained by the proprietary high-resolution material and shading system used by MetaHumans, which introduced some additional visual differences between the two avatar sets. Surface textures for the MetaHuman characters were selected manually within the MetaHuman Creator interface to approximate the appearance of the original GFG avatars. Directly transferring or replicating the exact GFG texture maps was not feasible due to differences in texture formats, resolutions, UV layouts, and the technical complexity of the MetaHuman material system. Furthermore, some features differ structurally between the two frameworks. For example, GFG eyebrows are implemented as a two-dimensional texture mapped onto the facial surface, whereas MetaHumans use strand-based systems with individually modelled three-dimensional hairs. Similarly, GFG characters do not include eyelashes, while MetaHumans include detailed eyelashes generated using the same strand-based hair system. These differences introduce additional minor visual variations between the avatar sets despite efforts to maintain close overall similarity (see *Fig.3.4*).

Finally, small residual differences may also have arisen from slight variations in virtual camera positioning and field of view across the two rendering pipelines, as well as differences in how lighting is implemented within the respective rendering engines. Given these technical constraints, the MetaHuman characters were matched to their GFG counterparts as closely as possible in terms of identity, appearance, and overall facial configuration. However, the combination of the previously described differences in mesh structure, materials, hair systems, and rendering inevitably results in characters that are similar in overall identity while still differing in certain visual details. These unavoidable technical differences are also reflected in the identity matching results reported later in this chapter where three of six identity pairs were not reliably judged as the same person.

Each avatar was presented in a 10-second video clip displaying extremely subtle, idle behaviours such as blinking, micro-saccades, and very minimal head sway. The same animation sequence was applied identically across all avatars to eliminate variability in movement dynamics. These animations were rendered at 60 frames per second and presented against a plain black background to maintain visual consistency and minimise contextual distractions.

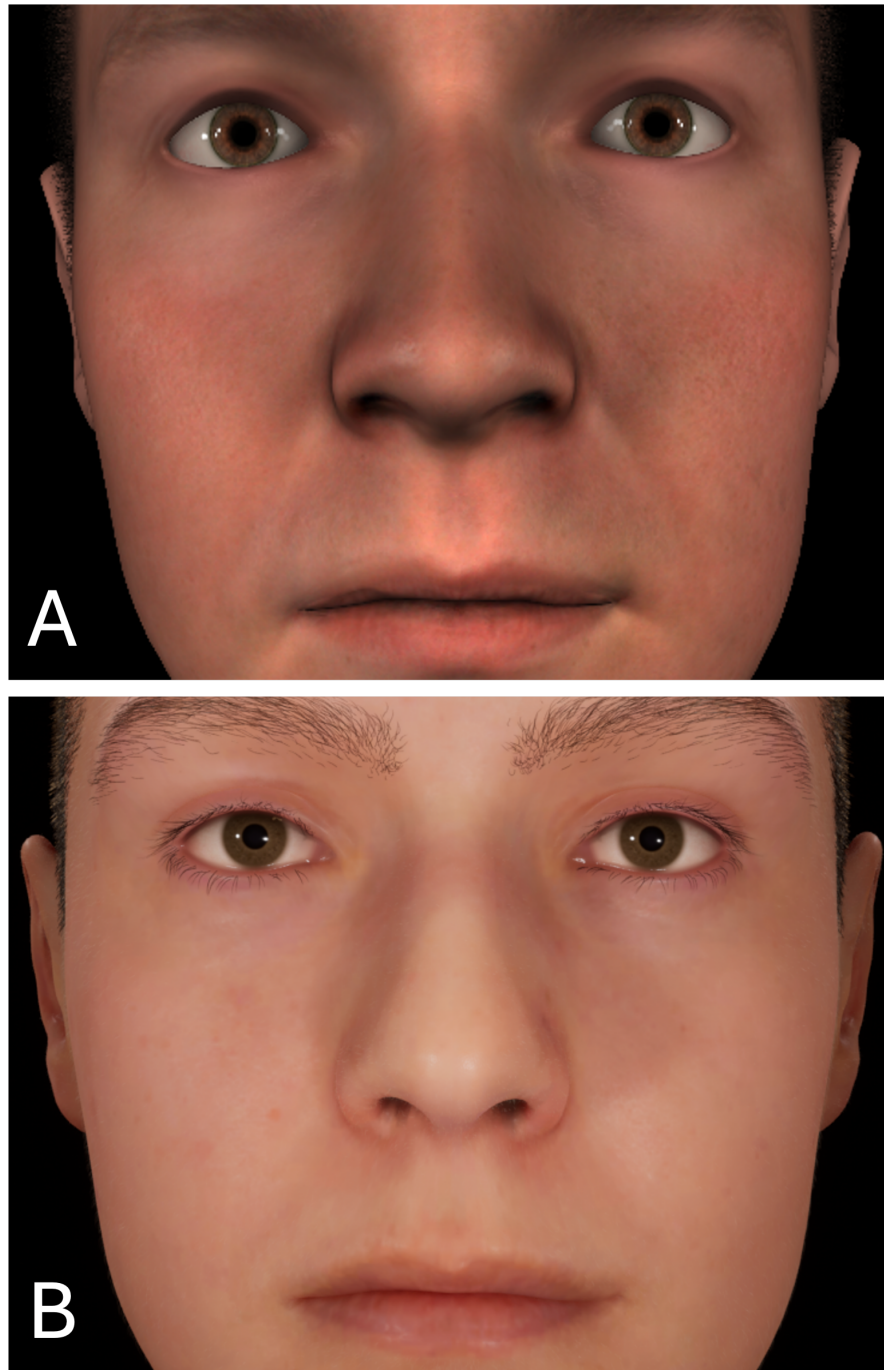


Figure 3.4: Visual comparison of surface detail between the original GFG avatar and its MetaHuman counterpart for the Caucasian male identity. (A) shows the GFG character, while (B) shows the reconstructed MetaHuman version generated using the Mesh to MetaHuman pipeline. Although the overall facial identity is preserved, several differences arise from the distinct material and hair systems used by the two frameworks. In addition to the previously described mesh topology differences, the GFG model represents eyebrows as a two-dimensional texture mapped onto the facial surface and does not include eyelashes. Whereas the MetaHuman character uses strand-based systems with individually modelled three-dimensional hairs for both eyebrows and eyelashes, along with higher resolution skin and eye materials. These differences contribute to the visual variations between the two avatar representations despite efforts to closely match the original identity.

3.1.2.4 Procedure

The experiment was administered online using a Qualtrics (Provo, UT) questionnaire. The study employed a within-subject (repeated-measures) design, where each participant viewed 12 short video clips (10 seconds in duration, 60 fps). To control for order effects, presentation order was counterbalanced across participants: half viewed all six GFG clips first, followed by the six MetaHuman clips, while the other half completed the reverse order. Within each block, the order of avatar presentation was also randomised for every participant. After each video, participants rated the avatar on eight bipolar seven-point Likert scales with a neutral midpoint (1 = low, 7 = high). The scales were drawn from validated measures by Ho and MacDorman (2010, 2017), designed to assess perceptual dimensions central to uncanny-valley research. Three items measured Humanness (Inanimate-Living, Synthetic-Real, Human-made-Humanlike), three measured Eeriness (Dull-Freaky, Predictable-Eerie, Plain-Weird), and two single-item scales measured Realism (Not realistic-Very realistic) and Familiarity (Strange-Familiar). Participants were instructed to rate each character according to their immediate subjective impression, with no time limit imposed.

Finally, demographic information was collected, and all responses were anonymised prior to analysis. For each participant, mean scores were computed per dimension by averaging ratings across all GFG or MetaHuman avatars separately, providing the dependent measures entered into subsequent statistical analyses.

3.1.3 Results

3.1.3.1 Uncanny Valley Measures

For each participant, mean scores were computed for each perceptual dimension (Humanness, Eeriness, Familiarity, and Realism) with higher scores indicating greater perceived humanness, eeriness, familiarity, and realism, respectively. To obtain framework-level values, ratings were averaged across all six avatars within each system (GFG and MetaHuman). For instance, a participant's GFG realism score represented the mean realism rating across all six GFG avatars, while their MetaHuman realism score represented the mean realism rating across all six MetaHuman avatars. This method produced one mean score per participant for each dimension and each avatar framework.

All statistical analyses were conducted in Python (v3.11; Van Rossum and Drake Jr, 1995) using the Pandas and SciPy stats packages. Data distributions met the assumptions for parametric testing and paired samples t-tests (two-tailed) were used to compare mean ratings between the GFG and MetaHuman avatars for each perceptual dimension. To control the family-wise er-

ror rate across the four paired comparisons, p-values were adjusted using the Holm-Bonferroni correction (Holm, 1979). The significance threshold was set at $\alpha = .05$.

The abbreviated human-likeness and eeriness composites demonstrated very high internal consistency, with Cronbach's $\alpha = .92$ and $\alpha = .89$, respectively. These values are comparable to those reported for the original full human-likeness ($\alpha = .87$) and eeriness ($\alpha = .86$) scales developed by Ho and MacDorman (2017). Internal consistency could not be assessed for the realism and familiarity measures as each of these constructs was assessed using a single item.

As shown in *Figure 3.5* and *Table 3.1*, participants rated the MetaHuman avatars as significantly more human-like ($M = 3.75$, $SD = 1.13$) than the GFG avatars ($M = 2.75$, $SD = 1.24$), $t(46) = -6.93$, $p_{Holm} < .001$. Eeriness ratings were significantly lower for MetaHumans ($M = 3.84$, $SD = 1.18$) than for GFG avatars ($M = 4.18$, $SD = 1.32$), $t(46) = 2.56$, $p_{Holm} = .014$. Familiarity ratings were higher for MetaHumans ($M = 3.98$, $SD = 1.31$) compared with GFG avatars ($M = 3.38$, $SD = 1.25$), $t(46) = -3.91$, $p_{Holm} < .001$. Finally, realism ratings followed the same pattern, with MetaHumans rated as significantly more realistic ($M = 3.99$, $SD = 1.14$) than GFG avatars ($M = 2.86$, $SD = 1.28$), $t(46) = -7.15$, $p_{Holm} < .001$.

Perceived Traits by Avatar Framework

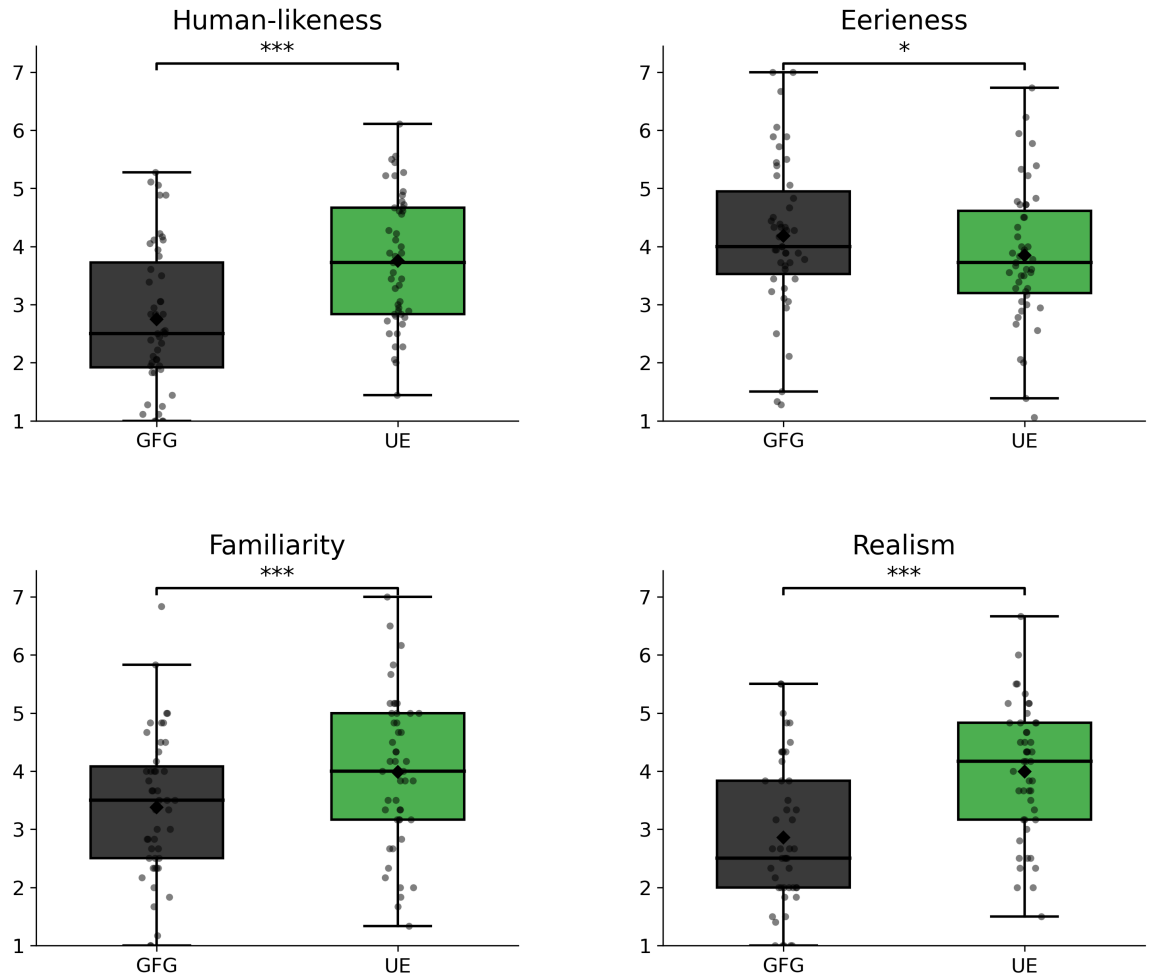


Figure 3.5: Boxplots showing participant ratings of avatars generated with the Generative Face Grammar (GFG) framework and Unreal Engine MetaHumans (UE) across four perceptual dimensions: human-likeness, eeriness, familiarity, and realism. The horizontal line inside each box shows the median rating, the diamond shows the mean rating, the box shows where the middle 50% of ratings fall, and the whiskers show the spread of most of the remaining ratings. Individual participant ratings are shown as transparent points. Ratings were made on a 7-point Likert scale. UE avatars were rated as significantly more human-like, familiar, and realistic than GFG avatars, whereas GFG avatars were rated as significantly less eerie. Asterisks indicate significance levels from paired samples *t*-tests following Holm-Bonferroni correction for multiple comparisons (* $p < .05$, *** $p < .001$).

	GFG			MH	
	M (SD)	$t(46)$	p_{Holm}	M (SD)	d_z
Human-likeness	2.75 (1.24)	-6.93	< .001	3.75 (1.13)	1.01
Eeriness	4.18 (1.32)	2.56	.014	3.84 (1.18)	0.37
Familiarity	3.38 (1.25)	-3.91	< .001	3.98 (1.31)	0.57
Realism	2.86 (1.28)	-7.15	< .001	3.99 (1.14)	1.04

Table 3.1: Mean ratings (M) and standard deviations (SD) for GFG and MetaHuman avatars across the four perceptual dimensions. Paired samples t -tests were used to compare ratings between the two avatar systems for human-likeness, eeriness, familiarity, and realism. Holm-adjusted p -values (Holm, 1979) and Cohen’s d_z (Lakens, 2013) effect sizes are reported.

3.1.3.2 Identity Matching

To assess whether the paired GFG and MetaHuman avatars were perceived as depicting the same individual identities, participants rated each of the six pairs on a 7-point similarity scale (1 = “Not at all the same person”, 7 = “Definitely the same person”). One-sample t -tests were conducted comparing the average similarity rating for each pair against the midpoint of the scale (4), which represents a neutral judgement indicating that the two avatars are perceived as neither clearly the same nor clearly different individuals. This approach allowed us to determine whether participants systematically perceived the paired avatars as depicting the same identity or as different individuals.

As shown in *Figure 3.6*, ratings were generally below the neutral midpoint, indicating that participants tended to perceive the GFG and MetaHuman versions as similar but not identical. The Caucasian male pair ($M = 2.77$, $SD = 1.94$) was rated significantly below the midpoint, $t(46) = -4.20$, $p < .001$, as was the Caucasian female pair ($M = 3.05$, $SD = 1.89$), $t(46) = -3.35$, $p < .01$, and the African male pair ($M = 2.56$, $SD = 1.78$), $t(46) = -5.32$, $p < .001$. In contrast, ratings for the Asian male pair ($M = 3.64$, $SD = 2.04$), $t(46) = -1.18$, $p = .24$, the Asian female pair ($M = 3.55$, $SD = 2.02$), $t(46) = -1.49$, $p = .14$, and the African female pair ($M = 3.45$, $SD = 1.95$), $t(46) = -1.86$, $p = .07$, did not differ significantly from the midpoint. Overall, ratings across most pairs fell slightly below the midpoint, indicating that the GFG and MetaHuman versions were generally perceived as similar but not identical.

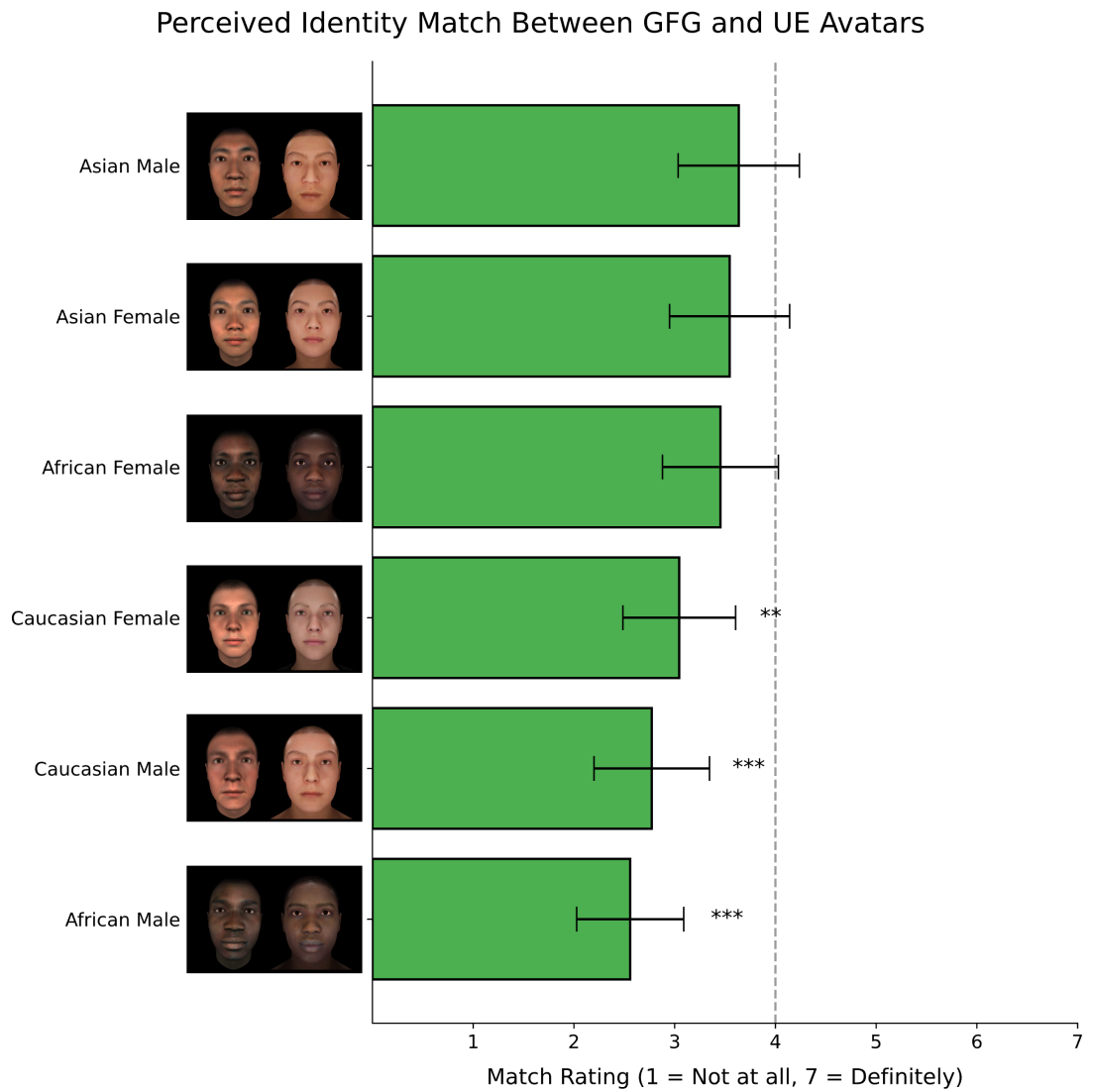


Figure 3.6: Perceived identity similarity between paired GFG and MetaHuman avatars for each character identity. Bars show the mean similarity rating on a 7-point scale (1 = “Not at all the same person”, 7 = “Definitely the same person”), with error bars representing 95% confidence intervals. The dashed vertical line indicates the neutral midpoint of the scale (4). Asterisks denote significant differences from the midpoint based on one-sample t-tests (** $p < .01$, *** $p < .001$).

3.1.4 Discussion

The present study examined whether highly fidelity MetaHuman digital avatars produced with Epic Games’ MetaHuman Creator (Epic Games, 2025) elevate perceived humanness, realism, and familiarity relative to established Generative Face Grammar (GFG) avatars (Yu et al., 2012), without eliciting stronger feelings of eeriness associated with the uncanny valley phenomenon (Kätsyri et al., 2015). The main aim was to evaluate whether the use of modern, high fidelity rendering technologies could improve the ecological validity of facial perception research, bringing

the visual quality of experimental stimuli closer to the standard of realism now common in digital media (Xu & Zhang, 2025).

While the GFG framework remains a pivotal tool for generating controlled, parameterised facial expressions and dynamic stimuli (Jack & Schyns, 2017), its visual rendering pipeline is quite dated by today's graphical standards, employing static illumination, limited skin texture rendering, limited hair rendering techniques, and low polygon counts in character topology. In contrast, MetaHuman Creator and the Unreal Engine 5 rendering framework integrate physically based rendering (PBR), subsurface scattering for realistic skin translucency, micro-normal maps capturing fine-grained surface detail, and dynamic global illumination that allows accurate light-surface interaction (Kaur & Josan, 2025). Together with advanced skeletal and facial rig systems capable of precise model deformations, the MetaHuman system enables the production of avatars that more closely approximate the visual appearance and dynamic behaviour of real human faces, particularly in terms of rendering fidelity and shading realism.

Across all measured dimensions, participants rated MetaHuman avatars as significantly more humanlike, realistic, and familiar than their GFG counterparts, while reporting significantly lower eeriness. The largest improvements were observed for perceived human-likeness and realism, whereas the reduction in eeriness, although statistically reliable, was comparatively more modest. This pattern suggests that improvements in visual fidelity can enhance perceptual realism without necessarily amplifying the uncanny valley related responses, demonstrating that MetaHumans are a good candidate for a successor to tools such as the GFG for use in psychological research.

Additional analysis examined whether the pairs of GFG and MetaHuman characters were perceived as depicting the same identities. Although similarity ratings indicated modest identity mismatches for some pairs, these differences likely reflect technical constraints of transferring characters between frameworks rather than genuine changes in facial identity. While the original GFG meshes were converted using the Mesh to MetaHuman workflow, the process requires the MetaHuman system to reconstruct missing anatomical and surface detail based on its internal facial model. Consequently, subtle differences in eyelid curvature, tear duct geometry, nasolabial folds, and surface reflectance may emerge during reconstruction. Moreover, GFG meshes rely on comparatively low-polygon geometry and simplified shading, whereas MetaHumans employ dense meshes with physically based materials and subsurface scattering along with high-fidelity hair rendering solutions. These rendering and reconstruction differences may therefore introduce small perceptual deviations in identity appearance without altering the underlying facial structure. Importantly, these variations are inherent to cross-framework character transfer and do not undermine the primary aim of the present study, which was to evaluate how modern rendering fidelity influences perceptions of humanness, realism, and eeriness.

The pattern of results observed in this study offers an important perspective within the context of the uncanny valley theory (Mori, 1970; 2012). According to Mori's original proposal, the relationship between human likeness and affective response follows a nonlinear curve, in which positive impressions increase alongside increases in realism up to a critical point, beyond which small imperfections in appearance (and/or motion) trigger sharp decline in positive affect, eliciting feelings of eeriness. More recent understanding of the phenomenon (Kätsyri et al., 2015) emphasises that these negative responses are not inevitable and likely depend on perceptual coherence; that is, whether the different visual cues of the stimulus remain consistent with its level of realism (e.g., realistic skin texture being consistent with realistic eye rendering).

3.1.4.1 Findings in Light of the Uncanny Valley Theory

The present findings can be interpreted in two ways in relation to the uncanny valley theory. In both interpretations of the results, the MetaHuman characters bring us closer to the appearance of real humans in comparison to the GFG characters without evidence of the sharp decline in affect or increase in eeriness predicted by the uncanny valley. Where the two potential interpretations come in is in relation to which side of the uncanny valley do the MetaHuman characters reside on. In one case, the MetaHuman characters are more humanlike, however, not so much that they would look almost but not quite human: placing them somewhere before the valley along the humanlikeness spectrum. The other, we postulate more plausible, is that the MetaHuman characters have surpassed the uncanny valley and are so consistent in the realism of the different visual cues that they do not elicit the uncanny response in human observers. One may expect the realism ratings to be generally higher for MetaHuman characters to if the second interpretation was to be true, however, we postulate that due to the “plain” and “default” look of the characters (relatively smooth skin texture, buzz-cut hairstyles, no facial hair, no wrinkles or blemishes on the skin) resulting from trying to match the appearance of the low fidelity GFG characters, the humanlikeness ratings somewhat suffered. Other possible explanation is that the simple, subtle animations used with the characters reduced the humanlike perceptions.

One way to directly address this ambiguity would be to design an experiment that systematically maps the nonlinear trajectory of the uncanny valley using the MetaHuman framework itself. Because MetaHuman characters allow precise manipulation of graphical fidelity through level-of-detail (LOD) settings, it would be possible to generate a realism continuum ranging from low-fidelity ‘stylised’ faces to highly detailed photorealistic characters while holding facial identity and animation constant. Participants could then rate humanness, eeriness, realism, and familiarity across this continuum. Such a design would allow researchers to empirically trace the shape of the human-likeness and affect relationship predicted by Mori (1970) and determine whether modern photorealistic avatars approach, enter, or surpass the hypothesised valley.

In fact, some studies which employed MetaHumans suggest that our second interpretation of the characters bridging the uncanny valley is also likely to be the case, and our reasoning behind somewhat lower than expected humanlikeness ratings could also be correct. Bae et al. (2025) presented participants with short videos featuring either less advanced and less realistic digital human characters or the MetaHumans and asked them to rate visual and behavioural humanlikeness, familiarity, social presence, perceived novelty, satisfaction, and purchase intention using scales derived from Ho and MacDorman's (2010) humanness index and related measures. MetaHumans were rated as more visually and behaviourally humanlike, more familiar, and more satisfying than the lower fidelity characters. Crucially, scatterplot analyses of human likeness versus familiarity suggested that digital humans clustered on the left side of the uncanny valley, whereas MetaHumans, at least in terms of appearance, clustered on the right side of the valley; behavioural human likeness, by contrast, still clustered on the left side of the valley for both character types.

There are, however, two caveats that slightly reduce how strongly this can be taken as evidence that MetaHumans have definitively overcome the uncanny valley. Firstly, Bae et al. operationalised affective response primarily via familiarity (alongside satisfaction and purchase intention), rather than more direct indices of eeriness or discomfort. As Diel et al.'s (2021) meta-analysis notes, familiarity-type measures can certainly be informative, but they are influenced by a broader mix of factors (e.g., warmth, aesthetics, social appeal) and do not always map cleanly onto uncanny valley-specific affect. For that reason, Diel et al. recommend using more targeted eeriness/negative-affect scales when drawing strong conclusions about uncanny valley effects, just like we have in the present study. This does not invalidate Bae et al.'s findings, however, it does mean that their claim that meta-humans have "overcome" the uncanny valley rests partly on an indirect measure.

Secondly, the human likeness contrast in Bae et al., was implemented primarily between technologies (digital human vs meta-human): within each category, realism was not parametrically manipulated. As Diel et al. (2021) also highlight, many studies that rely on coarse category comparisons rather than graded manipulations of visual fidelity make it difficult to pinpoint exactly where a given stimulus sits on the uncanny valley curve, hence why we cannot make this conclusion from our findings either. In Bae et al., the apparent placement of digital humans on the left and meta-humans on the right side of the valley is therefore inferred from average differences between two representative exemplars of each technology, rather than from a continuous sweep of realism levels within each class. Taken together, these design features mean that Bae et al.'s results provide converging, but not definitive, evidence that hyper-realistic MetaHuman characters can achieve higher human likeness and user acceptance without eliciting stronger uncanny feelings, a pattern that aligns closely with the present findings.

More direct support for this interpretation comes from Higgins et al. (2021), who examined

MetaHumans at different levels of graphical fidelity using Ho and MacDorman's (2010) human-ness and eeriness indices. They created high-fidelity MetaHumans (LOD0; dense geometry, detailed textures, strand-based hair, full facial rig) and lower-fidelity versions of the same characters (LOD4; reduced polygon count, simplified textures and hair, less complex rigging), and animated them with motion-captured positive and negative facial expressions. Here, LOD refers to "level of detail", which is a standard practice used in the videogames industry for optimisation of game assets. Therefore, the characters used were exactly the same, just with reduced graphical fidelity. Participants rated each clip on humanlikeness, eeriness, and attractiveness, as well as the intensity of expressed emotion. Across the board, high fidelity MetaHumans were judged to be more humanlike, more attractive, and less eerie than their lower fidelity counterparts, and were also perceived as expressing emotions more intensely. Higgins and colleagues therefore concluded that current state-of-the-art photorealistic MetaHumans "are no longer stuck in the uncanny valley", at least under the conditions tested. One potential explanation for the elevated eeriness of the lower LOD MetaHumans could be the cue mismatch between realistic facial expression animations and less realistic and detailed rendering of the characters (Kätsyri et al., 2015), although this was not tested directly.

From the perspective of Diel et al.'s (2021) recommendations, Higgins et al., (2021) addresses several of the limitations noted above: eeriness is measured directly rather than via familiarity alone, and realism is manipulated within a single character creation framework by altering level of detail, rather than by comparing entirely different avatar pipelines. At the same time, some important constraints remain. The study still contrasts only two fidelity levels rather than sampling a continuous realism gradient. Thus, while Higgins et al., provides stronger evidence than Bae et al., (2025) that highly detailed MetaHumans can be perceived as more humanlike and less eerie than lower-fidelity characters, it cannot precisely locate where these characters sit on the uncanny valley curve.

Taken together, however, Bae et al.'s (2025) large-scale survey data on familiarity and acceptance, Higgins et al.'s (2021) more tightly controlled within-MetaHuman manipulation of visual fidelity, and the present finding that MetaHumans are rated as more humanlike, more realistic, more familiar, and modestly less eerie than lower-fidelity GFG avatars all converge on the same conclusion: when implemented with coherent, high-quality visual cues, modern MetaHuman characters can move considerably closer to real-human appearance without triggering the sharp rise in eeriness predicted by classic formulations of the uncanny valley.

3.1.4.2 Implications and Positioning Within the Thesis

The implications of the present findings for experimental psychology and the broader thesis narrative can be summarised as follows.

What the present study establishes

- MetaHumans provide a high-fidelity stimulus platform for social perception research. The present results show that MetaHuman avatars are perceived as more humanlike, more realistic, and more familiar than GFG avatars without eliciting stronger feelings of eeriness.
- Experimental control can be retained alongside improved realism. Although legacy systems such as GFG and FaceGen have played a foundational role in enabling parametric control of facial expressions and identities (Dawel et al., 2022; Jack & Schyns, 2017; Todorov et al., 2011), MetaHumans allow researchers to maintain similar levels of experimental manipulation while substantially increasing graphical fidelity and ecological validity.
- Graphical fidelity may influence perception of emotional expressions. Previous work shows that identical motion-captured facial expressions appear more intense when rendered on higher-fidelity MetaHuman models (Higgins et al., 2021). This suggests that studies using lower-fidelity models may underestimate the perceptual richness of facial signals.

How this study fits within the thesis narrative

- **Study 1a (present study)** validates MetaHuman avatars as credible stimuli for experimental psychology and social perception research.
- **Study 1b** investigates how observers conceptually position MetaHumans relative to GFG avatars as “characters” or “people”, examining perceptions of emotional capacity and moral concern.
- **Study 2** shifts focus from appearance to behaviour by applying reverse correlation methods to identify motion patterns that observers perceive as natural in idle facial movement.
- **Study 3** directly tests how variations in idle face motion naturalness influence perceptions of humanness, realism, familiarity, and eeriness, thereby examining behavioural effects on uncanny valley responses.

Together, these studies aim to bridge technological realism with psychological realism, developing a unified understanding of how the visual and behavioural fidelity of digital humans shapes their perceived aliveness.

Directions for future research

- Systematically manipulate graphical fidelity across multiple levels of detail within the same MetaHuman identities to map the relationship between human likeness and affective responses.
- Independently manipulate appearance realism and motion realism to directly test mismatch-based accounts of the uncanny valley.
- Examine how specific visual components of realism (e.g., skin shading, eye rendering, hair simulation, facial motion dynamics) contribute to humanness and eeriness judgments.
- Investigate how observer characteristics such as cultural background or familiarity with digital humans influence responses to highly realistic avatars.

3.1.5 Conclusion

Overall, this study shows that MetaHuman avatars can surpass older tools such as GFG in perceived humanness, realism, and familiarity, while also producing a modest reduction in perceived eeriness rather than an increase, positioning them as a robust and ecologically valid stimuli for experimental work on social perception. At the same time, the finding that MetaHumans are seen as more humanlike and person-like than GFG avatars raises an important question that quantitative ratings alone cannot fully answer: how do people make sense of these characters, and in what ways do they see them as more humanlike?

The next study (*Study 1b*) therefore turns to a qualitative, mixed-methods approach, using open-ended reflections and personhood-related ratings to examine how participants spontaneously describe, interpret, and morally position MetaHuman versus GFG characters subjectively.

3.2 Study 1b: Qualitative Evaluation of MetaHuman and GFG Avatars

3.2.1 Study Overview

As digital humans and avatars become more common in everyday settings, from entertainment and games to education, and online social platforms, the key question is no longer whether people notice them, but how they perceive and relate to them. When a character appears lifelike, observers begin to form social impressions. Observers begin to form judgements about whether it appears capable of feeling or whether it has a clear personality. These impressions matter because they influence trust, empathy, engagement, and co-presence, all of which are central to effective human-avatar interactions (Blascovich et al., 2002; de Melo et al., 2014; Kim et al., 2020; Nowak & Fox, 2018; Schroeder & Adesope, 2014; Waytz, Cacioppo, & Epley, 2010). For designers and researchers, the implications are both practical and theoretical, as social attributions can shape whether people are willing to collaborate with virtual agents, accept guidance or feedback from them, and extend moral concern in situations involving harm or fairness.

As described in *Study 1a*, the present chapter uses a single, multi-stage online questionnaire to examine how differences in avatar appearance relate to both perceptual evaluations and socio-psychological judgements. *Study 1a* focused on perceptual evaluations of the avatars, including human likeness, realism, familiarity, and eeriness, providing insight into how observers experience the visual qualities of characters that differ in graphical fidelity. The present study (*Study 1b*) examines a complementary set of social and cognitive judgements collected within the same questionnaire, including perceived capacity for emotion, perceived personality, person versus object framing, and moral concern. In addition to these quantitative measures, *Study 1b* also includes open-ended questions that capture participants' subjective reflections on the characters, allowing for qualitative insight into how and why particular impressions of the characters may be formed. As both studies (*1a* and *1b*) use the same stimuli and were completed by the same participants, this approach allows for a direct comparison of two avatar frameworks, GFG and MetaHuman, across perceptual and higher-level social dimensions.

Thus, allowing the present chapter to address a key question: do advances in appearance fidelity simply change what people see, or do they also shape how people relate to and judge digital characters? By pairing perceptual measures (*Study 1a*) with social and moral judgements (*Study 1b*) within the same context, the chapter examines whether differences in visual realism and human likeness are accompanied by inherent differences in psychological and moral attributions. This design establishes a clear progression from *Study 1a's* perceptual findings to *Study 1b's* socio-psychological focus.

3.2.2 Introduction

3.2.2.1 Continuity from Study 1a

Study 1a compared two frameworks for generating parametrically controlled digital avatars, the GFG (Yu et al., 2012) and Unreal Engine’s MetaHuman system (Epic Games, 2021), and identified clear perceptual differences between them. Across all key measures, MetaHuman avatars were rated as more humanlike, more realistic, and more familiar than their GFG counterparts, whilst also rating the MetaHumans as significantly less eerie than the GFG characters. These results point to perceptual advantages associated with more modern rendering approaches, likely driven by improvements in lighting, materials, and surface detail.

Building directly on these results, the present study (*Study 1b*) examines whether such framework differences extend beyond perceptual realism to influence socio-cognitive judgments. Specifically, examining whether participants attribute greater emotional capacity, personality, personhood, and moral concern to avatars created in the MetaHuman framework compared with those generated in the GFG. As both sets of ratings were collected within the same questionnaire and from the same participants, we can examine whether the perceptual distinctions observed in *Study 1a* correspond to meaningful differences in how participants relate to these avatars socially. Thus, the remainder of this literature review broadens the lens: from how real avatars look to how human they feel.

Research suggests that increases in graphical fidelity may do more than simply enhance visual believability; they can also influence the social inferences people make about digital agents. When avatars display more humanlike visual cues, observers may be more likely to interpret them in terms of agency and emotional capacity, drawing on general mechanisms of mind perception rather than responding specifically to visual fidelity alone (H. M. Gray et al., 2007; Martini et al., 2016; Waytz, Cacioppo, & Epley, 2010). However, this work does not demonstrate a direct causal relationship between graphical fidelity and anthropomorphism; instead, it shows that humanlike appearance cues are often used by observers to infer whether an agent is capable of intention and experience. Such appearance-based inferences are central to anthropomorphic perception and can shape social judgments towards digital humans, including perceptions of their capacity for empathy or trustworthiness. Thus, the following section turns to anthropomorphism as a conceptual framework for understanding how differences in visual fidelity may influence social and moral attributions toward virtual characters.

3.2.2.2 Introducing Anthropomorphism

Anthropomorphism refers to the tendency to attribute human-like mental states, such as thoughts or emotions, to non-human entities (Epley et al., 2007). This tendency helps people interpret artificial or ambiguous agents by relying on the same social understanding they use when interacting with real humans. In the context of digital avatars, anthropomorphism provides the psychological basis for perceiving characters as emotionally capable or deserving of moral consideration.

A related line of work on mind perception differentiates between attributions of agency (the capacity to act intentionally) and experience (the capacity to feel) as two core dimensions underlying perceptions of humanness (Gray et al., 2007). These dimensions play a key role in moral judgement: seeing an entity as capable of feeling leads to empathy and moral concern, while seeing it as capable of intentional action leads to judgments about responsibility and blame (K. Gray et al., 2012). Although this framework is not specific to avatars, it provides a theoretical basis for understanding how perceived mental capacities in artificial agents can shape empathetic and moral responses. While qualitative work shows that human-like and realistic avatar designs are associated with higher attributions of agency and emotional capacity (Mazerant, Schouten, et al., 2025), experimental research demonstrates that increases in human-like appearance influence mind attribution in artificial agents, with observers showing a categorical shift toward treating agents as intentional once a certain appearance threshold is reached (Martini et al., 2016).

Importantly, this literature does not suggest a simple direct relationship between graphical fidelity and anthropomorphism. Instead, it indicates that observers rely on human-like appearance cues, such as facial structure and eye-related information, when deciding whether to interpret an entity in terms of having a mind (Looser & Wheatley, 2010; Waytz, Gray, et al., 2010). This cue-based account is further supported by recent digital avatar research examining a wide range of avatar forms. Across large sets of avatars spanning human, humanoid, and non-human designs, human-like sensory features such as eyes, mouths, and skin were more strongly associated with attributions of agency and experience than motoric features such as limbs (Mazerant, van Berlo, et al., 2025). This suggests that facial (and other sensory) cues play a key role in mind attribution, even when overall avatar form varies. However, this work does not manipulate the differences in realism within human avatars; thus, not addressing the question of whether similar cue-based mechanisms operate when comparing less and more realistic digital humans. Qualitative research also implies that people rely on cues such as facial detail and character realism when deciding whether an avatar can think or feel (Mazerant, Schouten, et al., 2025). The researchers suggest that increasing visual fidelity does not simply lead to a gradual increase in perceived mindedness. Mind attribution, again, seems to depend on whether an avatar meets a

perceptual threshold defined by key human-like cues, after which it is more likely to be seen as capable of thought and feeling.

Thus, in digital human research, anthropomorphism appears to function at the intersection of perceptual and social realism. While high-fidelity rendering may shape how real a character appears, anthropomorphic attributions reflect whether that character is perceived as having a mind ‘behind the face’. The present study builds directly on this distinction by examining whether the perceptual advantages observed for MetaHuman avatars in *Study 1a* are accompanied by stronger attributions of emotional capacity, personality, personhood, and moral concern. Having established anthropomorphism as a guiding framework, we will now examine how differences in visual fidelity may potentially translate into these socio-psychological judgements associated with anthropomorphism.

3.2.2.3 From Realism to Emotional and Moral Attribution

As digital human technology evolves, one important question is whether making a character look more real also changes how viewers respond to them socially and morally. The uncanny valley literature suggests that highly human-like appearances can elicit discomfort, but not as an immediate consequence of increased character realism (Kätsyri et al., 2015; Mori, 1970; Mori et al., 2012). More specifically, negative reactions are most consistently observed when human-like features are perceptually inconsistent or mismatched; for example, when realistic facial detail is paired with unnatural motion, rather than when character realism simply increases. Based on the mismatch account of the uncanny valley, visual realism alone does not determine social or affective responses, but it does seem to interact with processing related to perceptual coherence.

Alongside perceptual accounts, mind-perception theories propose that human-like agents may lead to attributions of mental capacity, particularly the capacity to feel and experience, which in turn shape how those agents are evaluated socially and morally (K. Gray & Wegner, 2012). However, recent work has questioned whether mind attribution by itself can explain uncanny responses. Through a meta-regression analysis and a longitudinal ‘(de)humanisation’ experiment, MacDorman (2024) demonstrated that attributing experience consistently increased eeriness in vignette-based studies, but that these effects were significantly reduced or non-existent when observers were presented with visual stimuli directly. In these cases, physical appearance and perceptual factors appeared to have a stronger influence on eeriness than mind attribution itself. Crucially, these findings do not undermine the role of mind perception in social evaluation more generally. They highlight a distinction between early, more stimulus-driven perceptual responses such as feelings of eeriness and later, more reflective judgements about an agent’s emotional capacity and moral standing. This suggests that mind attribution may play a limited role in immediate uncanny reactions while playing a key role in higher-level social and

moral evaluations of digital human characters.

Higgins et al. (2023) tested whether visual fidelity alone changes people's reactions to virtual humans by showing participants MetaHuman characters at two levels of detail: a high-realism version (LOD0) and a reduced-realism version (LOD5). Across a series of short video clips, participants rated state empathy, broken into cognitive (understanding how the character feels), affective (feeling what the character feels), and compassion (concern/sympathy) measures, alongside common uncanny valley measures. Their results showed that visual realism did not produce a main effect on any of the anthropomorphic empathy dimensions, and it also did not have a main effect on eeriness or familiarity. Critically, the authors pointed out that the lack of significant effects of realism on uncanny valley measures was surprising and contradictory given that their earlier study using the same characters and similar uncanny valley measures did show a significant effect of realism (Higgins et al., 2021).

They suggested that one reason for this inconsistency may be the experimental design: in the 2023 paper, visual realism was manipulated between-subjects (each participant saw either high realism or low realism, but not both), whereas the 2021 study employed a within-subject design (the same participant sees both levels of realism). A within-subjects design can be more sensitive because participants have a direct reference point; seeing both versions can make differences more salient. Thus, the absence of realism effects in 2023 may partly reflect the lower sensitivity of a between-subjects design for detecting these effects. This could potentially also apply to the more “anthropomorphic” judgements like empathy or mind-related attributions. Given all these considerations, adopting a within-subjects design could provide a more sensitive test of whether changes in graphical fidelity influence higher-level socio-cognitive/anthropomorphic attributions when other variables are held constant.

Overall, this body of work suggests that visual realism alone may not determine social or moral judgements of digital characters. These judgements are thought to depend on the presence of specific human-like cues that support mind attribution. Accordingly, the present study combines quantitative anthropomorphic ratings with qualitative reports of the cues participants focus on when explaining perceived differences between avatar frameworks whilst employing a repeated-measures design.

3.2.2.4 Visual and Perceptual Features Underlying Anthropomorphism

Building on cue-based accounts of digital character perception, prior work suggests that viewers do not respond to realism as a single universal property. Instead, impressions of digital human characters are formed by integrating multiple visual and behavioural cues, including surface appearance, rendering style, motion quality, and the overall coherence between these features (Kätsyri et al., 2015; MacDorman et al., 2009; McDonnell et al., 2012). When these cues align,

characters tend to be judged as more convincing and appealing. When they conflict, observers frequently report a vague sense that something feels off, a response commonly associated with perceptual mismatch explanations of uncanny impressions.

On the appearance side, photorealistic facial replication is consistently associated with higher judgements of human-likeness in digital characters (Bae et al., 2025; Fraser et al., 2024; McDonnell et al., 2012). Examinations of facial modelling and animation techniques identify surface detail and anatomical structure as core components of realistic digital faces, including features such as wrinkles, skin texture, and illumination consistency (Kaur & Josan, 2025; Xu & Zhang, 2025). In virtual reality (VR) social interaction contexts, avatars with greater human resemblance and greater graphical resolution are more often judged as realistic and preferred, particularly when compared directly with less detailed alternatives (Fraser et al., 2024). As emphasised by the aforementioned research, closely related cues include anatomical plausibility and surface detail: faces that respect human structural constraints and exhibit realistic skin characteristics are less likely to draw attention to artificiality than faces containing subtle irregularities in structure or texture (MacDorman et al., 2009).

Rendering choices further shape how these appearance cues are perceived. Lighting and rendering, including shading, shadows, and material response, influence whether facial surfaces behave in a physically plausible way and can reduce artefacts that undermine perceived realism (McDonnell et al., 2012). Studies where fidelity manipulations are applied to MetaHuman characters also tend to alter hair representation alongside other visual features, and these various changes have been linked to differences in overall evaluations such as human-likeness, attractiveness, and eeriness (Higgins et al., 2021), even though individual features such as hair are not typically isolated as single causal drivers (Higgins et al., 2023), they likely play a role in overall impressions of the characters.

Dynamic cues play an equally important role in how observers evaluate digital human characters. Facial animation quality, including coordinated motion across facial regions and subtle micro-movements, contributes to whether a character comes across as convincing (Higgins et al., 2023; McDonnell et al., 2012). More broadly, disruptions to expressive facial dynamics, such as reduced movement in the upper face during speech, can amplify uncanny impressions even when other aspects of the character are coherent (Tinwell et al., 2011). Eye behaviour is particularly salient in this respect. Natural blinking patterns and eye movements are often identified as signals of vitality and responsiveness, especially during periods of inactivity (Fu et al., 2008; Ruhland, Peters, et al., 2015). Work on interaction with avatar-based virtual agents shows that the absence of such behaviours can draw attention to artificiality, whereas low-amplitude idle motion in the eyes and head helps maintain a sense of presence (Ruhland, Zibrek, & McDonnell, 2015). However, even small irregularities in eye timing or gaze coordination are noticeable and can undermine the overall credibility of an otherwise realistic character (Fu et al., 2008).

Beyond realism and naturalness, these perceptual cues can also induce emotional engagement. In immersive VR, virtual character design has been associated with variation in users' reported 'feelings of being', including engagement, trust, enjoyment, effectiveness, and security. Character preferences show moderate correlations with some of these feelings, suggesting that character type and presentation can meaningfully influence how socially and affectively involving an experience becomes (Dirin & Laine, 2023). When cues are poorly integrated, viewers often describe a lingering sense of unease or "wrongness," consistent with perceptual mismatch accounts, motivating the treatment of reports such as "something feels off" as a meaningful descriptive category in qualitative data (Kätsyri et al., 2015; MacDorman et al., 2009).

These literature-derived cues informed the deductive coding framework used in the present qualitative analysis (see the *methods* section). Specifically, participant responses were coded with respect to five theory driven perceptual categories: surface realism, structural accuracy, dynamic expressivity, emotional and social presence, and perceptual incoherence. Together, these dimensions capture how visual and behavioural cues are thought to support anthropomorphic perception and social evaluation of digital human characters.

3.2.2.5 Research Aims and Hypotheses

The present study builds directly on the findings of Study 1a and on cue-based accounts of anthropomorphism. *Study 1a* showed that MetaHuman avatars were judged as more humanlike, more realistic, more familiar, and less eerie than GFG avatars. These perceptual differences suggest that the two frameworks differ in some ways across specific human-like cues and not simply in visual realism alone. Prior work indicates that such cues play a central role in whether observers attribute mental states, personality, and moral relevance to artificial agents (Epley et al., 2007; Gray et al., 2007; Waytz et al., 2010; Kätsyri et al., 2015).

Thus, the first aim of the present study is to examine whether the perceptual advantages observed for MetaHuman avatars extend beyond appearance-based evaluations to higher-level social and moral judgements using quantitative methods. Participants rate both GFG and MetaHuman avatars on four social-cognitive attributes central to anthropomorphism: perceived emotional capacity, perceived personality, person-like versus object-like perception, and moral concern. Based on the perceptual results of *Study 1a* and on literature linking human-like cues to mind attribution and moral concern, we test the following hypothesis: **H1:** MetaHuman avatars will receive higher ratings of emotional capacity, perceived personality, personhood, and moral concern than GFG avatars.

The second aim of the study is qualitative and explanatory. Instead of testing a particular hypothesis, this component aims to explore what cues participants use when forming their judgements about the avatars across the two frameworks. Through open-ended responses, participants

describe whether they perceived differences between GFG and MetaHuman avatars and which features shaped their impressions of humanness, emotional presence, or person-likeness. These reports are analysed using a literature-informed coding framework, allowing participants' explanations to be linked to specific perceptual cues such as facial detail or the quality of rendering.

Together, these aims reflect a mixed-methods approach. The quantitative analysis tests whether framework-level differences in social and moral attribution emerge, while the qualitative analysis clarifies why those differences may occur by revealing the cues participants themselves rely on. The following section details the methods used to implement this design.

3.2.3 Methods

3.2.3.1 Participants

All data for *Study 1b* were collected as part of the same online questionnaire used in *Study 1a*. Following completion of the main section (which included the perceptual realism and uncanny valley measures reported in *Study 1a*), participants were given the option to continue with an additional section containing further questions. This optional section forms the basis of the present study. As compensation, participants who chose to complete this optional section received an additional entry into a prize draw for one of two £50 Amazon vouchers.

As described in *Study 1a*, recruitment was conducted via leaflet advertisements across the University of Glasgow and word-of-mouth opportunity sampling. Eligibility criteria required participants to be aged 18 or older, have normal or corrected-to-normal vision, and to report no (1) lexical, reading, language, or hearing atypicality (e.g., dyslexia); (2) social or emotional processing atypicality (e.g., autism spectrum disorder, depression, anxiety); or (3) face-processing atypicality (e.g., prosopagnosia).

A total of 44 participants (21 male, 23 female) out of the original 47 completed this portion of the study, with a mean age of 30.52 years ($SD = 10.27$). All participants met the inclusion criteria. Ethical approval for the research was granted by the University of Glasgow College of Medical, Veterinary & Life Sciences Ethics Committee (Ref: 200240259).

3.2.3.2 Measures

Socio-cognitive and anthropomorphic judgements of each avatar were assessed using a combination of quantitative rating scales and open-ended qualitative questions. All measures were administered within the same online questionnaire as *Study 1a* and were completed after the perceptual realism ratings.

Participants first rated each avatar on four statements assessing key dimensions of social and psychological attribution, using 5-point Likert scales (1 = ‘*Strongly Disagree*’, 5 = ‘*Strongly Agree*’). Specifically, across the four statements, they indicated the extent to which they agreed that: (1) “*this avatar appears capable of feeling emotions*”; (2) “*this avatar seems to have a distinct personality*”; (3) “*this avatar feels more like a person than an object*”; and (4) “*I would feel uncomfortable if this avatar were mistreated*”. These items captured perceived emotional capacity, personality, person-likeness, and moral concern, respectively; these concepts are central to anthropomorphic and socio-cognitive engagement with artificial agents (H. M. Gray et al., 2007; Waytz, Cacioppo, & Epley, 2010). Research in human-robot interaction shows that anthropomorphic design affects expectations of agency, emotional capacity, and appropriate social and moral treatment, supporting the relevance of such judgements for evaluating human responses to artificial agents (Złotowski et al., 2015). Prior experiential work also supports the idea that observing the mistreatment of robots elicits negative affect and empathic concern, validating discomfort with harm as an indicator of social engagement (Rosenthal-von der Pütten et al., 2013). For each participant, mean scores were computed per each of the four dimensions, with higher scores indicating stronger agreement with the attribution in question. Because each socio-cognitive dimension was assessed using a single item, internal consistency statistics were not applicable for these measures.

Each quantitative item was conceptually derived from established explanations of anthropomorphism and mind perception (H. M. Gray et al., 2007; Waytz, Cacioppo, & Epley, 2010). The item assessing emotional capacity reflects the experience dimension of mind perception, which captures the attribution of feelings and subjective states. The personality and person-likeness items index broader social attributions commonly associated with agency and person-oriented framing in anthropomorphism research. The moral concern item assesses the extent to which an avatar is viewed as deserving of care and consideration, particularly in contexts involving harm or mistreatment, consistent with mind perception theory and prior human-robot interaction research (H. M. Gray et al., 2007; Rosenthal-von der Pütten et al., 2013; Złotowski et al., 2015).

All participants also answered a single question indicating which pronoun felt most natural to use when referring to each avatar (“he”, “she”, “it”, “they”, or “other”). This item was included as an exploratory behavioural measure of social categorisation, indexing the extent to which avatars were framed as person-like versus object-like and, where appropriate, gendered. Prior work on linguistic anthropomorphism suggests that pronoun selection reflects implicit social framing rather than explicit beliefs about an agent’s mental capacities (Abercrombie et al., 2021; Nowak & Fox, 2018).

In addition to the aforementioned quantitative measures, participants provided written responses to three open-ended questions designed to capture what influenced their impressions

of the avatars. They were asked whether they found themselves thinking differently about the two avatar types (GFG versus MetaHuman), what visual or motion features made the avatars seem as if they had a personality or emotional presence, and which aspects of physical appearance (e.g., skin, eyes) made the avatars seem more or less human-like. These responses were thematically analysed using the deductive coding framework outlined in *Table 3.2*, linking participants' observations to cues of anthropomorphic perception (see the *analysis* section below for more details).

Note: an additional quantitative question in the optional portion of the questionnaire asked participants to rate the extent to which each GFG-MetaHuman pair appeared to represent the same person (“*To what extent do the two avatars look like they could represent the same person?*”; 1 = ‘*Not at All*’, 7 = ‘*Definitely*’). However, because this question concerned perceived identity similarity rather than socio-cognitive attribution, it was analysed alongside the perceptual realism data in *Study 1a* and is reported in that portion of this chapter.

3.2.3.3 Stimuli

The stimuli used in *Study 1b* were identical to those described in *Study 1a*, consisting of six pairs of digital avatars originally created in the GFG framework and their corresponding transfers into Unreal Engine’s MetaHuman system. As detailed in *Study 1a*, each GFG mesh was converted into a MetaHuman using the Mesh to MetaHuman workflow and manually adjusted in MetaHuman Creator to match the appearance of the original GFG character in terms of skin tone, eye colour, and facial structure. Each pair therefore represented the same virtual identity rendered through two distinct graphical pipelines, differing predominantly in texturing, lighting, and material fidelity due to each framework’s internal rendering solutions.

For the present study we used still images of the characters. Each GFG-MetaHuman pair was displayed side by side in a single frame, with the GFG avatar shown on the left (labelled ‘A’) and its MetaHuman counterpart on the right (labelled ‘B’). This configuration allowed participants to make direct visual comparisons between the two representations of the same character (see *Fig.3.7* for example characters). In the questionnaire, under each image pair, participants responded to the anthropomorphism-related questions described in the measures section. These questions were answered separately for each avatar (A and B) presented in the given pair, allowing direct comparison between frameworks while controlling for facial identity. The pronoun attribution question was presented in the same format.

For the open-ended qualitative questions, participants were shown a single composite image displaying all avatars together. On the left side of the image, all six GFG characters were presented with the label ‘A’, and on the right side all six MetaHuman characters were shown with the label ‘B’ (see *Fig.3.7*). This layout encouraged participants to reflect on broader framework-

level differences rather than individual identities of the characters.



Figure 3.7: Stimulus characters used in the experiment. Six identities were first generated using the Generative Face Grammar (GFG) framework and then transferred to the MetaHuman system using the Mesh to MetaHuman tool in Unreal Engine. (A) shows the original GFG avatars, while (B) shows the corresponding MetaHuman counterparts reconstructed from the GFG meshes. Each row represents a matched identity pair, with three male and three female characters spanning three broad ethnic groups (Caucasian, East Asian, and African). The MetaHuman characters were subsequently refined in MetaHuman Creator by manually selecting the closest matching hair style and adjusting facial and eye textures to approximate the appearance of the original GFG identities as closely as possible.

3.2.3.4 Procedure

The data for *Study 1b* was collected within the same online questionnaire used in Study 1a. After completing the initial section containing the perceptual realism and uncanny valley related measures, participants were invited to take part in an optional section of the questionnaire. This optional section contained the additional socio-cognitive measures and open-ended questions

described in the measures section. Participants who chose to complete this portion received an additional entry into a prize draw for one of two £50 Amazon vouchers as an incentive to participate.

The structure and general procedure mirrored that of *Study 1a*. Participants first viewed the still image pairs depicting the GFG (A) and MetaHuman (B) versions of each identity, presented side by side on screen. For each pair, they completed the four socio-cognitive rating items and the pronoun attribution question. Once all six avatar pairs have been rated, participants answered the open-ended qualitative questions using a text box field. Participants were instructed to base their qualitative reflections on their overall impressions of the two groups rather than any specific individual character/identity.

Participants were encouraged to provide their immediate, intuitive impressions when completing the rating items and to elaborate freely in the open-ended text boxes. No time limits were imposed for either section. As in *Study 1a*, responses were anonymised, and basic demographic information (age and gender) was recorded at the start of the questionnaire.

3.2.3.5 Analysis

Quantitative

Quantitative analysis focused on participants' Likert scale ratings of the four socio-cognitive dimensions: emotional capacity, personality, personhood, and moral concern, as well as their pronoun selections for each avatar. For each participant, mean scores were computed separately for GFG and MetaHuman avatars across all six characters per framework. Higher mean values reflected stronger agreement that an avatar possessed emotional or personality characteristics, felt more person-like than object-like, or elicited greater moral concern if mistreated. Repeated measures t-tests were conducted to compare GFG and MetaHuman avatars on each of the four socio-cognitive dimensions, using two-tailed tests with a significance threshold of .05. Descriptive statistics (means, standard deviations) and inferential results (*t* and *p* values) were reported for each comparison.

Responses to the pronoun attribution question were plotted as frequency distributions across categories (he, she, it, they, other). This data provided an additional insight into perceived personhood and gender attribution. Gender comparisons between frameworks were summarised descriptively to highlight general patterns of linguistic anthropomorphism.

All quantitative analyses were carried out using Python (version 3.11; Van Rossum & Drake Jr, 1995) using the pandas, NumPy, and SciPy packages.

Qualitative

The qualitative analysis focused on the three open-ended questions focusing on assessing participants' subjective assessment of the avatars across the two frameworks. Responses were analysed using a deductive thematic approach (Braun & Clarke, 2006) guided by a coding framework derived from prior literature on anthropomorphism and perceptual realism in digital humans (see *Table 3.2*). The framework included five codes reflecting important perceptual cues of anthropomorphic response. This approach ensured that qualitative analysis remained theoretically grounded while allowing inductive elaboration of subthemes as new patterns emerged within participants' descriptions.

All text responses were imported and analysed into NVivo 14 (Dhakal, 2022). Initial coding was performed line by line, with each extract assigned to one or more relevant codes. Coding included both features that participants explicitly mentioned and features inferred from broader evaluative statements, even when specific visual or behavioural cues were not explicitly mentioned. Following initial coding, related categories were grouped into higher-level themes describing how specific perceptual cues potentially informed the different socio-cognitive attributions of the different avatar frameworks. To make the analysis more thorough, the coding framework and emergent themes were reviewed and discussed with a second researcher; any disagreements were resolved through discussion, and the framework was refined to ensure that interpretations remained theoretically grounded and faithful to the original data. Illustrative quotations were selected to represent each major theme and to ground interpretations in participants' own language.

Code	Description	Supporting Literature
Surface Realism	Visual realism of the face’s surface and materials (skin texture, pores, wrinkles, shading, lighting, subsurface scattering, reflections, hair detail)	(Bae et al., 2025; Fraser et al., 2024; H. M. Gray et al., 2007; MacDorman et al., 2009; McDonnell et al., 2012; Waytz, Cacioppo, & Epley, 2010)
Structural Accuracy	The underlying anatomy and proportions of the face & head (facial geometry, bone/muscle placement, eyelid or mouth shape, skull structure, alignment of features)	(MacDorman et al., 2009; McDonnell et al., 2012)
Dynamic Expressivity	Perceived movement and emotional responsiveness (facial animation, micro-expression, blinks, gaze behaviour, subtle motion suggesting thought or feeling)	(Fu et al., 2008; Higgins et al., 2023; McDonnell et al., 2012)
Emotional & Social Presence	The sense that the avatar feels alive or capable of emotion/personality (comments about empathy, warmth, engagement, or perceived mind/personhood)	(Dirin & Laine, 2023)
Perceptual Incoherence	When realism cues conflict or something feels “off” (mismatch between appearance and movement, plasticity, lifelessness, categorisation difficulty)	(Kätsyri et al., 2015; MacDorman et al., 2009)

Table 3.2: *Deductive coding framework outlining key visual and perceptual features associated with anthropomorphism in digital humans, derived from prior literature. The codes were used to guide the thematic analysis of participant responses in the present study, with accompanying descriptions and supporting references for each category.*

3.2.4 Results

3.2.4.1 Quantitative

As described in the analysis section, mean ratings were calculated for each of the four socio-cognitive dimensions (emotional capacity, distinct personality, person-likeness, and moral con-

cern) separately for GFG and MetaHuman avatars. For each participant, scores were averaged across the six avatars within each framework, such that higher values indicated stronger agreement that an avatar possessed emotional or personality characteristics, felt more person-like than object-like, or evoked greater moral concern if mistreated.

A series of repeated measures t-tests compared GFG and MetaHuman avatars on each of the four dimensions. Participants rated MetaHuman avatars significantly higher than GFG avatars on emotional capacity, $t(43) = 2.63, p = .012$; person-likeness, $t(43) = 4.07, p < .001$; and moral concern, $t(43) = 3.59, p = .001$. Ratings of distinct personality were not significantly different between the two frameworks, $t(43) = 1.18, p = .244$. See Fig.3.8.

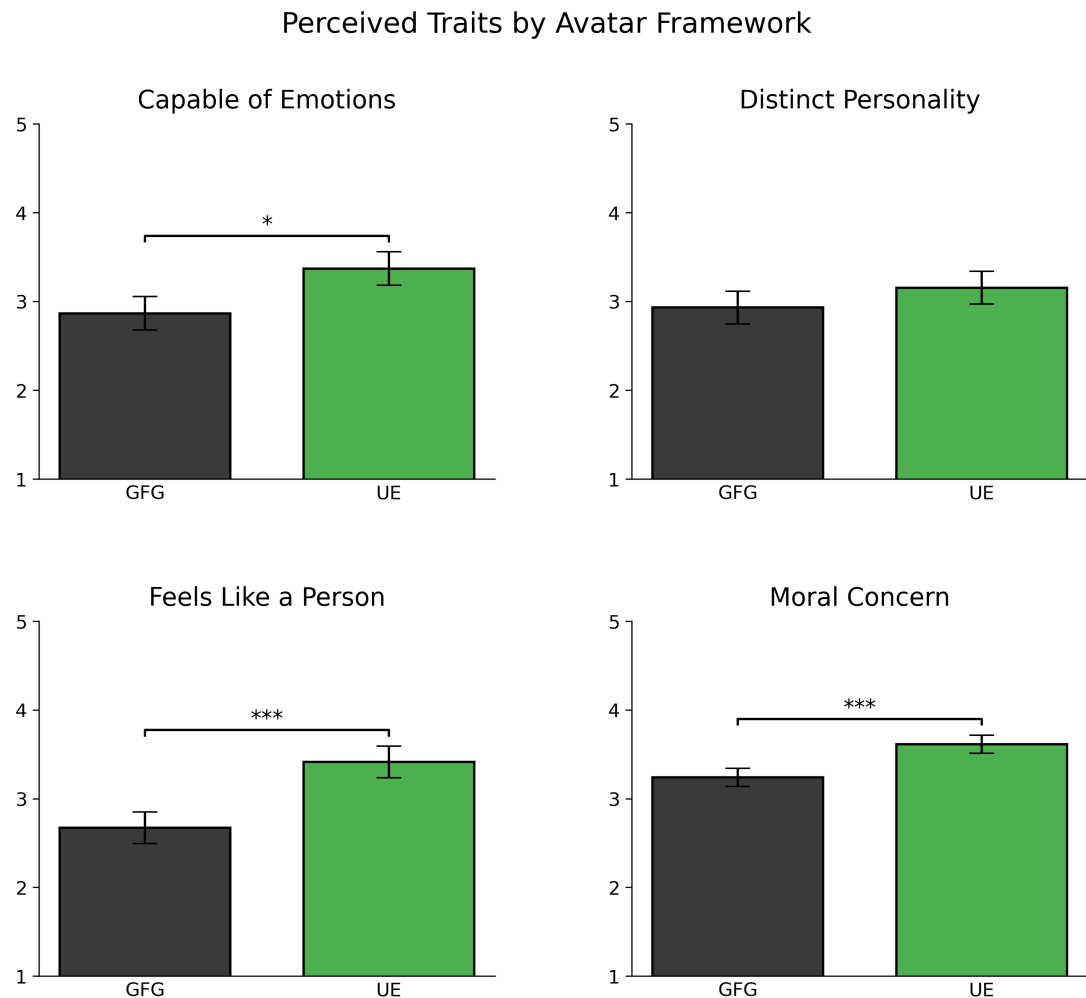


Figure 3.8: Mean ratings of socio-cognitive attributes for GFG and MetaHuman (UE) avatars across four dimensions: emotional capacity, distinct personality, person-likeness, and moral concern. Ratings were made on a 5-point scale, with higher values indicating stronger agreement that the avatar possessed the given attribute. Error bars represent ± 1 standard error of the mean. Asterisks indicate significant differences between frameworks (* $p < .05$, *** $p < .001$). MetaHuman avatars were rated significantly higher than GFG avatars on emotional capacity, person-likeness, and moral concern, while no significant difference was observed for distinct personality.

Pronoun attribution was included as an exploratory, supplemental measure to provide additional insight into how avatars were socially categorised, rather than as a primary outcome of the study (see *Fig.3.9*). Across both avatar frameworks, pronoun use generally followed perceived gender cues, indicating that participants tended to frame the avatars as person-like rather than object-like.

For GFG avatars, pronoun use was relatively variable. Male avatars were predominantly referred to as “he”, whereas female avatars showed mixed patterns, with substantial use of “she” alongside “he”, “they”, and “it”, suggesting greater ambiguity in gender attribution.

In contrast, the MetaHuman avatars elicited more consistent use of gendered pronouns. Female MetaHuman avatars were referred to primarily as “she”, with markedly reduced use of alternative or non-gendered pronouns compared to their GFG counterparts, while male MetaHuman avatars were overwhelmingly referred to as “he” (except for ‘African Male’ MetaHuman, but even then, the attribution was very dominated by the use of “she”).

Use of “it” and “they” was relatively infrequent overall but appeared to occur more often for GFG avatars than for MetaHumans. These patterns were consistent across avatar identities, indicating that framework-level differences in visual realism and rendering, rather than identity-specific features, influenced pronoun attribution.

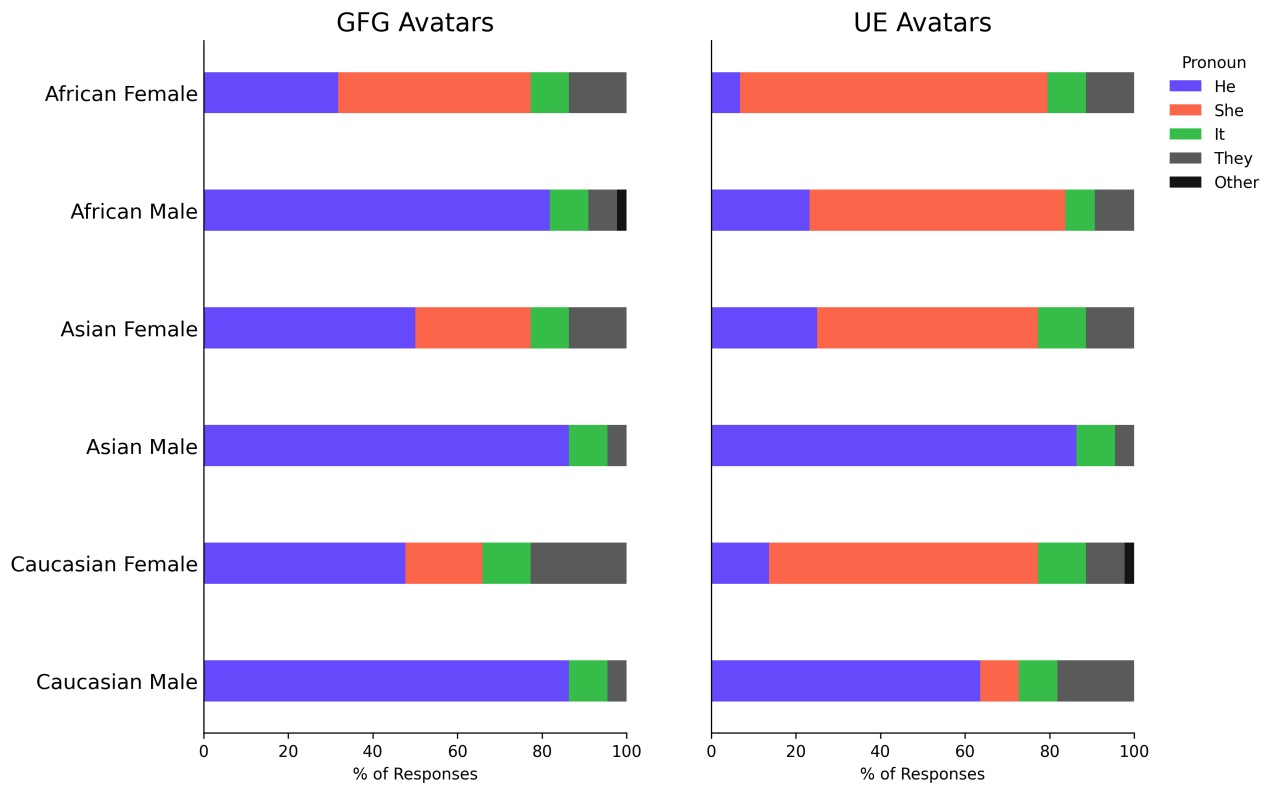


Figure 3.9: *Distribution of pronoun attributions (%) for GFG and MetaHuman (UE) avatars across character identities. Stacked bars represent the proportion of responses assigned to each pronoun category (he, she, it, they, other) for each avatar. Across both frameworks, pronoun use generally followed perceived gender cues; however, MetaHuman avatars elicited more consistent use of gendered pronouns, whereas GFG avatars showed greater variability and more frequent use of non-gendered or objectifying terms (e.g., “it” and “they”).*

3.2.4.2 Qualitative

Codes derived from the initial framework (Table 3.2) and from inductive framework refinement were grouped into higher-level themes reflecting different aspects of participant reasoning. Some themes captured evaluations of visual appearance and perceptual coherence, while others reflected more general or interpretative judgements about inner life and character depth.

Theme 1: Character Appearance

This theme captures how participants used static facial cues to judge how humanlike, real, or artificial the avatars appeared. Across responses, impressions were strongly anchored in facial structure (e.g., proportions and feature placement) and surface rendering quality (e.g., skin texture, colour blending, and lighting). These cues were often treated as the most immediate basis for deciding whether an avatar looked like a “real person” versus a stylised or synthetic

character, and they shaped broader reactions such as comfort, trust, and ease of relating.

A consistent pattern was that MetaHuman avatars were described as more humanlike and realistic overall, whereas GFG avatars were more often framed as game-like, generic, or visually “off”. Many participants highlighted the eyes as the most influential facial feature, describing the GFG avatars as having eyes that were “*too wide and too white*”, which reduced realism as explicitly stated by one of the participants: “*the big eyes in A [GFG] made them seem less realistic*”. On the other hand, MetaHumans were more often described as having eyes with greater realism thanks to subtle details which helped the faces come across as more lifelike. More specifically, as one of the participants worded it: “*For me I noticed that I felt a character was more lifelike if the eyes were more realistic - most notably colour or subtle redness in the eyes*”. Another participant highlighting that “*Eyes and mouths. These enhanced most of B [MetaHuman] and detracted from most of A [GFG]. Some other aspects were easier to overlook [...] but the eyes and mouths were most important*”. Despite these dominating views on the eyes of the avatars across the two frameworks, a single participant noted that “*the biggest knock against all of them was their eyes [...]. And you shouldn't see the white under the iris*”; thus overall, positioning the eyes as a critical cue driving perceptions of realism and artificiality.

Alongside the eyes, participants frequently referred to skin and facial surface detail as a cue that influenced how they perceived the avatars across the two frameworks. MetaHumans were often described as having more convincing skin thanks to superior textures, shading, and more natural colour blending, while GFG avatars were often characterised as overly simplified. One participant specified that “*The skin texture is essential to making them look real. Group A [GFG] feels like 20 years older in terms of graphics*”, explicitly linking one of the key differences in how the characters are perceived to the rendering quality of the skin. Others made similar comparisons in favour of the MetaHumans through mentions of wrinkles, blemishes, uneven tone, and the way lighting fell across facial features as cues that made an avatar feel closer to real human skin. In general, participants’ views are broadly reflected in this quote by one of the participants: “*[...] eyes of Group B [MetaHuman] added to the overall realism as did the texturing of skin, and lips. The skin colours also seemed much more naturally blended.*”, although, one participant noted that “*B [Metahuman] seems more consistent and too perfect without any blemishes. On the other hand A [GFG] had way too many differences especially in terms of colour on the face*”. Overall, many participants stated that skin rendering played a big role in how the avatars were perceived and, in most cases, the MetaHuman skin rendering was seen as superior in terms of driving character realism and human-likeness.

Participants also used structural language to describe whether faces looked anatomically plausible and human-like. Several comments pointed to proportions that felt wrong or exaggerated in the GFG characters beyond the previously mentioned “*wide-eyed and staring*” look. One participant mentioned that “*The characters in category A [GFG] feel a little bit more eerie than*

those in category B [MetaHuman]. It's difficult to pinpoint exactly what is different, but the facial features feel a little bit softer in category B", another said that "B [MetaHuman] just seemed more human in terms of their facial features". Even when participants did not pinpoint specific structural features, they still described facial configuration as influencing how humanlike the characters appeared, with a general trend towards MetaHumans being seen as more humanlike in terms of their facial features. For a more explicit facial structure comment, one participant noted that "*the Type B [MetaHuman] avatars presented much more human-like features and so it's easier to relate to them, while the Type A [GFG] avatars feel much more generic [...] one characteristic that is very distinct between type A [GFG] and B [MetaHumans] is their noses*" with a handful of participants also explicitly mentioning the cheeks, lips and foreheads influencing their judgements. Overall, participants' responses indicate that perceptions of structural accuracy were grounded in holistic impressions of facial configuration rather than precise identification of individual features. These global judgements influenced whether faces were experienced as anatomically plausible and human-like, with MetaHuman avatars more consistently described as structurally coherent and easier to interpret as human characters.

Although mentioned by relatively few participants, a small subset offered more technical reflections on how camera perspective and facial presentation influenced perceived humanlikeness for them. These responses moved beyond individual features to focus on how the face was visually framed and situated in 3D space, particularly in the GFG avatars. One participant described the camera perspective as face distorting, stating that "*Camera perspective on A [GFG] feels so much bizarre. Almost like the fish-eye lens seems to squeeze their face, causing a lack of cheeks*", while another similarly explained that "*I feel like the camera 'lens' pointing toward them made a difference to me. The more I could see of their face (aka the less of a 'fish eye' sorta view there is), the more I see them as humanlike*". This sense of distortion was also linked to the impression that facial features were over emphasised, compromising overall head structure, with one participant noting that "*there is an emphasis on facial features, as though that is all that needs to be communicated, with the rest of the head pushed back out of sight*". On the other hand, realism was described as increasing when facial features appeared integrated within a 'coherent three-dimensional head' rather than exaggerated in isolation. As one participant expressed, "*eyes that had depth of color and physical embeddedness in the head read as more human*", suggesting that human-likeness depended not only on feature detail but on how salient those features were in the 3D space. Taken together, these less common but nuanced observations indicate that perceived realism was shaped not only by surface texture or anatomical accuracy, but also by more technical factors such as viewpoint, camera lens, and the spatial coherence of the face as a whole.

The last appearance-related pattern involved participants explicitly framing the avatars in a videogame context. Several participants drew a clear distinction between the two frame-

works, with one stating that “*Avatars A [GFG] look like they’ve been taken out of a video game whereas avatars B [MetaHumans] look like real life people*”, and another similarly noting that “*B [MetaHumans] seemed more human like, A [GFG] seems more video game like*”. This interpretation was elaborated on in one of the more detailed responses where one participant compared the GFG characters directly to a specific videogame aesthetic, stating: “*A [GFG] reminds me strongly of the video game Oblivion [a videogame released in 2006] with strange saturation points, odd shapes [...]. There is an emphasis on facial features, as though that is all that needs to be communicated, with the rest of the head pushed back out of sight*”. In this account, exaggerated facial emphasis and wide-eyed expressions were not framed as faults, but as a form of stylised characterisation, with the participant noting that “*the facial expressions themselves are wide-eyed and staring, which gives them a kind of personality in the same way that a cartoon has personality*”. On the other hand, the MetaHumans were described as attempting realism through improved surface rendering and less exaggerated proportions: “*B [MetaHumans] appears to be trying for realism, the skin textures are better [...] and the eyes less wide and manic*” making it “*easier to ascribe a surface-level assumption of race and sex to these avatars*” and they came across as “*capable of greater nuance of expression, largely through the better textures and more lifelike ‘default’ expressions and not because of any change in movement*”. Although this participant concluded that “*none of the avatars read as ‘human’ to me but B certainly reads as closer to human than A*”. This response illustrates how videogame associations shaped participants’ interpretations of appearance, sometimes positioning stylisation as almost intentional rather than simply unrealistic or artificial. Taken together, these accounts suggest that videogame references functioned as a broader interpretive framework through which some participants evaluated character appearance, influencing whether visual exaggeration was read as unrealistic or simply belonging to a different representational category.

Taken together, these accounts show that judgements of character appearance were grounded in a combination of lower-level visual cues (such as eye shape, skin texture, and facial proportions) and higher-level interpretations of how coherently those cues were integrated together to form a plausible human face. Instead of simply focusing on individual features in isolation, the participants had a tendency to evaluate whether facial elements worked together to produce a coherent face that felt anatomically believable. MetaHuman avatars were more often described as meeting these criteria, while GFG avatars were more frequently interpreted as stylised, game-like, or visually exaggerated, even when such stylisation was occasionally experienced as a positive feature. Importantly, these appearance-based judgements were typically surface-level, providing an initial ‘filter’ through which avatars were classified as more or less humanlike. As the following theme shows, however, participants did not rely on appearance alone: evaluations of inner life, expressivity, and psychological presence often emerged once participants elaborated on aspects beyond static facial cues to consider what the characters seemed capable of feeling, thinking, or expressing.

More broadly, these qualitative findings also point to a possible route for future quantitative work. The perceptual cues identified here, such as eye-region realism, skin rendering, facial proportions, and structural coherence, could be isolated and manipulated systematically in controlled experiments to test their individual and combined contribution to perceived humanness. For example, specific properties of the eye region could be varied while holding all other features constant, allowing direct assessment of how features such as sclera detail, iris size, or eyelid shape influence perception. In practical terms, modern character systems such as MetaHuman provide highly customisable eye materials, enabling fine-grained manipulation of these properties within a consistent facial model. In this way, the present themes could form the basis of a more quantitative framework for refining digital avatars towards more human-like perception.

Theme 2: Signs of Inner Life

This theme captures how participants evaluated whether the avatars appeared capable of emotion, awareness, or psychological presence, moving beyond static appearance to consider whether the characters felt ‘alive’ in a social or experiential sense. Participants did not focus solely on facial realism; they described making judgements about inner life based on subtle cues suggesting intentionality or emotional capacity. These evaluations were often expressed in interpersonal terms, such as feeling empathy, trust, or connection towards an avatar, and were not always associated with surface realism. In many cases, avatars described as more visually realistic were simultaneously experienced as blank, or emotionally distant, while less realistic avatars were sometimes perceived as more expressive or relatable. Across responses, signs of inner life were inferred from subtle facial cues: particularly around the eyes and the perceived capacity of the face to convey expression, suggesting that perceived aliveness depended less on surface realism and more on impressions of psychological presence.

Participants frequently described emotional and social presence in terms of whether an avatar felt capable of eliciting empathy or trust. Importantly, these judgements were not consistently aligned with visual realism. While some participants reported that the more realistic MetaHuman avatars felt easier to empathise with, others experienced these same characters as emotionally distant and blank. For example, one participant noted that “*Avatars A [GFG] look less realistic but are more appealing to me/seem like they have more personality. Avatars B [MetaHumans] seem more realistic but uncanny and generic*”, while another similarly characterised the MetaHumans as “*more realistic but [...] emotionally distant*”, describing them as having a “*blank stare*”. In this sense, increased realism did not consistently translate into emotional engagement and, in some cases, appeared to elevate feelings of eeriness or emotional detachment.

At the same time, some participants reported stronger emotional engagement with the GFG

avatars, despite these characters being generally perceived as less visually realistic overall across participant responses. One participant stated that they “*felt more emphatic towards A [GFG] cause they are more alive for me*”, while another explained that “*Type A [GFG] avatars looked more real, so I found myself more convinced they’re showing a certain emotion*”. Some other responses aligned with this pattern, suggesting that emotional presence was sometimes attributed even when participants acknowledged stylisation or artificiality of the GFG characters. For example, one participant described the GFG avatars as having “*emotional eyes*” whereas the MetaHuman avatars were characterised as having a “*blank stare*”.

However, this pattern was not universal across the responses. Many participants instead described stronger feelings of trust or comfort towards the MetaHuman avatars, attributing emotional and social presence to subtler facial cues and overall ‘softness’ rather than expressiveness. One participant noted that “*the characters in category A [GFG] feel a little bit more eerie than those in category B [MetaHuman]*”, adding that they “*would trust those in category B [MetaHuman] more*” even though they found it difficult to articulate why. In this account, emotional engagement was linked to perceived facial softness, colouration, and shading, with the participant highlighting that “*the facial features feel a little bit softer in category B [MetaHuman]*”, and that “*the colours in category B [MetaHuman] are also more humanlike, particularly the lips and the shadows on the face*”. Other participants also emphasised the feelings of relatability and comfort, describing the GFG avatars as feeling “*less real, more robotic and simulated, you feel less of a human connection to them*”. Similarly, another participant contrasted the two avatar frameworks by noting that “*Type B [MetaHuman] avatars presented much more human-like features and so it’s easier to relate to them, while Type A [GFG] avatars feel much more generic and static*”, further supporting the idea that emotional presence was sometimes grounded in perceived naturalness and approachability rather than potential for expressiveness some saw in the GFG characters. Taken together, these accounts show that emotional and social presence was judged using multiple, sometimes conflicting criteria, with participants differing in whether they prioritised potential for expressiveness, realism, or interpersonal comfort when deciding which avatars felt more human or relatable.

Some participant responses also specifically referred to character motion when describing signs of inner life. As described in the methods section of *Study 1a*, during the initial portion of the questionnaire participants saw short, animated clips of all the characters. Every avatar was animated using the exact same short and subtle motion sequence, consisting of very slight head sways, tiny eye movements, and blinks. Thus, references to differences in movement across avatars cannot be attributed to differences in the animation itself but instead reflect how these identical motions were visually implemented in each framework; for example, through differences in 3D model deformation such as movement of the eyelids or the surrounding eye region.

Some participants treated the subtle dynamics of the eye region as important indicators of whether an avatar came across as alive or as possessing inner life. Judgements of emotional presence were often linked to blinking and small eye movements. For instance, one participant who described the GFG characters as *“less real, more robotic and simulated”* and that *“you feel less of a human connection to them”*, also emphasised the importance of *“movement around the eyes, not just the eyes blinking in isolation but more subtle facial movements under and to the side of the eyes and around the nose”* which is a feature of the more detailed deformation of the eye region in the MetaHuman characters. Another participant expressed responding generally positively to *“unexpected blinking and what I perceived as tiny eye orbital movements”*. These evaluations were often described as operating independently of, or alongside, surface appearance, suggesting that perceived inner life was shaped by how facial dynamics were interpreted rather than by visual realism alone.

When participants explicitly compared the two avatar frameworks in terms of dynamic expressivity, there was no general consensus, instead the responses described different ways of interpreting subtle motion. Several participants described the MetaHuman avatars as exhibiting more natural or convincing dynamics, particularly around the eyes and blinking. These avatars were often described as benefiting from *“less exaggerated movements”*, with one participant noting that MetaHumans *“seemed more natural to me, despite being less expressive”*, and another reporting that *“the blink response was more natural in B [MetaHumans], but still not totally human like”*. In these accounts, the motion of the MetaHumans was interpreted as supporting realism and human-likeness due to being smoother despite the fact that both avatar frameworks used the exact same animation curves. However, some participants differed in how they interpreted these dynamics across the two frameworks. Some experienced the MetaHumans’ motion as emotionally flat or static, reporting that this reduced their sense of aliveness. Whereas, the GFG avatars were sometimes described as feeling more expressive or emotional, indicating that the perception of inner life was not interpreted the same across all participants.

Taken together, the findings from this theme indicate that participants did not rely on a single cue when judging whether avatars appeared to possess inner life. The characters’ emotional and psychological presence was inferred through a combination of subtle facial dynamics, perceived expressiveness, and subjective impressions of trust or comfort. Visual realism alone did not appear to guarantee emotional connection: more realistic avatars were sometimes experienced as emotionally distant or unsettling, while less realistic avatars were occasionally perceived as more expressive or relatable. These responses emphasise the individual differences in how inner life was attributed, suggesting that perceived aliveness was influenced more strongly by impressions of psychological presence than by surface realism.

Theme 3: Depth and the Uncanny

Participants often described a general sense of depth or inner presence in the avatars, noting that its absence made characters feel uncanny or unsettling. Rather than pointing to specific facial features or clear emotional cues, participants described an overall sense that an avatar either felt internally coherent or hollow. These thoughts and feelings were often difficult to articulate and were portrayed almost as intuitive reactions to the avatar as a whole. This theme differs from the preceding ones by capturing a holistic, hard to explain judgement of the avatars, reflecting participants' difficulty in explaining why an avatar felt convincing or unsettling without referencing specific visual elements.

When participants attempted to describe what contributed to this sense of depth or inner presence, their explanations often focused on the eye region but remained vague and metaphorical. Several participants referred to the eyes as fundamental to whether an avatar felt lifelike, not simply in terms of realism but in relation to a perceived 'depth' beyond simple appearance. One participant noted that a character felt *"more lifelike if the eyes were more realistic – most notably colour or subtle redness in the eyes and depth"*, while another similarly described avatars as feeling *"more real by the eyes and depth of their skin"* adding that *"the depth of the eyes too [...] contributes a lot"*. These accounts often blurred the line between physical description and subjective impression, with 'depth' referring both to the way the eyes were situated in the face and the depth of the eyes in a broader, almost intangible sense. In more affective terms, this was expressed through metaphors of inner life, such as when one participant stated that *"the avatars in the B [MetaHuman] category had a soul"*, whereas the GFG characters *"felt like the lights were on but nobody is home"*. Others explicitly struggled to articulate the reasons for such impressions, responding that they *"cannot pinpoint what it is"*, or describing intangible cues in metaphorical language, such as *"looking up with a little brightness in the eye"*. In some cases, this sense of depth was framed as a feeling of being seen or engaged by an actual being rather than a simple 3D computer model representing a person, with one participant explaining that *"something in the expression made some of them seem real. Like they were looking at me and not over my head or right through me"*. Together, these responses suggest that perceived depth was experienced as an emergent, intuitive quality, often closely related to the eyes; however, this was frequently articulated as a feeling of internal presence rather than a clearly identifiable visual feature.

When participants compared the two avatar frameworks, impressions of depth and uncanniness were often mixed and sometimes contradictory. Several participants described the MetaHuman avatars as more human-like and closer to realism, with one stating that *"B [MetaHumans] look like much more human-like character models, and feel less within the realm of 'uncanny valley'"*. However, this perceived realism did not always prevent the eerie impressions. Some participants reported that although MetaHumans appeared more human, they simultaneously

felt emotionally vacant or unsettling, noting that *“the eyes on the first top 4 avatars of group B [MetaHuman] look hollow, almost like dead people”*, or that *“Group B [MetaHumans] seemed more human like than Group A [GFG], but equally as creepy looking”*. Others described MetaHumans as *“more realistic, but more dull”*, comparing this with the GFG avatars, which were seen as *“more eerie and not lifelike”*, yet sometimes more appealing. These conflicting views were also reflected in comments such as *“Avatars A [GFG] look less realistic but are more appealing to me/seem like they have more personality. Avatars B [MetaHumans] seem more realistic but uncanny and generic”*, suggesting that realism and perceived inner presence did not always align. In some cases, uncanniness was linked to specific features or mismatches, such as *“lack of eyelashes and empty eyes”* in the GFG avatars, while MetaHumans were described as benefiting from more detailed rendering, with one participant phrasing it as *“the expressions were better rendered in B [MetaHumans]. More detail = greater verisimilitude”*. Taken together, these responses suggest that uncanny impressions often occurred not due to a lack of realism specifically, but as a result of perceptual incoherence: instances where human-like appearance and perceived inner presence or expressiveness failed to align in a way that felt psychologically consistent.

Overall, this theme shows that perceptions of depth and uncanniness were shaped by more than surface realism or specific visual features. Participant responses suggest a more general sensitivity to whether an avatar appeared as coherent, whether its appearance and implied inner presence felt integrated in a way that supported the impression of a thinking or feeling entity. When this sense of coherence was disrupted, avatars appeared to be experienced as hollow or eerie, even when, visually, they came across as human-like. On the other hand, characters that lacked realism in physical appearance were sometimes perceived as more present or engaging if they conveyed a stronger impression of ‘inner substance’. These findings suggest that uncanny responses emerged from a complex perceived mismatch between outward human likeness and an inferred absence of inner life, posing ‘depth’ as an emergent, subjective quality rather than a property related to a specific aspect of appearance.

Summary of the Themes

Overall, participants’ impressions of the avatars across the two frameworks were driven by a combination of perceptual and evaluative factors. Many responses revolved around static visual appearance, including facial structure, eye configuration, and skin rendering, which influenced judgements of human likeness. Participants also inferred ‘inner life’ in the characters based on subtle facial cues and dynamic signals, viewing these as indicators of the potential for emotional and psychological engagement. In addition, many participants described a more abstract sense of depth or internal coherence, often expressed as subjective impressions rather than specific visual descriptions, which influenced whether avatars felt convincing or uncanny. Across these factors

contributing to the overall impressions of the avatars, there was noticeable individual variation, with participants weighing different cues in different ways. Together, the qualitative findings indicate that perception of digital humans consists of layered evaluations of appearance, expressivity, and inferred inner presence, highlighting the complexity of how human-likeness and uncanniness of digital humans are experienced. *Table 3.3* provides a summary of the prevalence and frequency of these themes and their associated subcodes across participant responses.

Theme	Code	Participants	Mentions
(1) Character Appearance	Structural Accuracy	19	25
	Surface Realism	38	67
	Technical References	2	2
	Videogame References	4	5
(2) Signs of Inner Life	Dynamic Expressivity	27	41
	Emotional & Social Presence	19	20
(3) Depth and the Uncanny	Intangible & Depth	10	13
	Perceptual Incoherence	10	12

Table 3.3: Summary of qualitative themes and subcodes identified across participant responses. For each code, 'Participants' column indicates the number of individuals who referenced that feature at least once, while 'Mentions' column reflects the total number of coded instances across all responses. This summary provides an overview of both the prevalence and frequency of perceptual cues underlying participants' judgements of the avatars across the three main themes.

3.2.5 Discussion

The present study examined whether increased graphical fidelity influenced how digital avatars are both, perceived visually, and how they are evaluated socially and morally. *Study 1a* demonstrated that the higher graphical fidelity MetaHuman avatars were perceived as significantly closer to real human faces and significantly less unsettling than their lower fidelity GFG counterparts. Building on these findings, *Study 1b* investigated whether these perceptual differences were also reflected in socio-cognitive attributions. In the quantitative portion of the present study, MetaHuman avatars were rated higher than GFG avatars on emotional capacity, person (vs object) likeness, and moral concern, while no significant difference emerged for perceived personality.

Qualitative findings provided further insight into how these judgements may have formed, revealing that participants relied on complex and layered evaluations involving static appearance, subtle cues to expressivity, and a broader sense of perceived depth or internal coherence.

Together, the mixed-methods findings indicate that differences in graphical fidelity extended beyond surface realism and were linked to how participants interpreted the avatars' social and psychological qualities.

3.2.5.1 Socio-Cognitive Judgements

The higher ratings for MetaHuman avatars on emotional capacity, person (vs object) likeness, and moral concern suggest that differences in graphical fidelity shaped how participants attributed mind and moral standing to the characters. These effects indicate that the perceptual advantages observed in *Study 1a*, namely that MetaHuman avatars were judged as more human-like, more realistic, more familiar, and less eerie, were not limited to visual judgements but were associated with differences in socio-cognitive interpretation.

This pattern of results aligns with cue-based accounts of anthropomorphism and mind perception. Mind perception involves two core dimensions, agency and experience (Gray et al., 2007), with the emotional capacity item mapping onto experience, and moral concern reflecting a consequence of the attribution of the capacity to feel (K. Gray et al., 2012). Higher ratings for MetaHumans on the socio-cognitive judgements therefore suggest that improved human-like cues enhanced experience attribution. Importantly, this does not imply a simple linear effect of realism as observers appear to rely on specific facial cues, such as eye-region information, structural plausibility, and character coherence, when deciding whether an entity possesses a mind (Looser & Wheatley, 2010; Mazerant, van Berlo, et al., 2025; Waytz, Gray, et al., 2010), a pattern evident in the qualitative findings.

The results are also compatible with threshold models of mind attribution, which propose that sufficiently human-like cues can drive categorical shifts in inference of intention and emotional capacity (Martini et al., 2016). Although the present study did not directly manipulate perceptual thresholds, the higher ratings on socio-cognitive judgements for MetaHumans suggest that this avatar framework may provide the configuration of cues necessary to support mind attribution more often. At the same time, in line with MacDorman (2024), these findings indicate that higher-level social and moral evaluations can operate alongside the more immediate perceptual responses such as eeriness, rather than simply reflecting them. Together, the evidence suggests that perceptual coherence can support experience attribution and moral concern without treating realism as the sole determinant of anthropomorphic impressions.

Notably, no significant difference emerged for perceived personality. This null effect may reflect the relatively constrained design of the stimuli: avatars were presented with standardised hairstyles and neutral expressions, potentially limiting the availability of individuating cues typically used in personality judgements. Research indicates that personality inferences from faces can depend on subtle variations in internal facial features and their configuration (Kramer

& Ward, 2010). In this case, such cues may have been insufficiently differentiated across the two avatar frameworks to produce systematic personality differences, despite shifts in broader mind-related attributions.

3.2.5.2 Subjective Impressions of Avatars

The qualitative findings shed some light on the processes underlying the quantitative differences observed between the two avatar frameworks. Instead of simply responding to graphical fidelity as a single global property, participants instead appeared to integrate multiple cues when forming impressions about the human-likeness and inner life of the characters. Static visual features such as eye shape, skin texture, shading, and facial structure often served as the basis for initial classification. MetaHuman avatars were frequently described as more realistic and anatomically coherent, whereas the GFG avatars were more likely to be characterised as stylised or with exaggerated proportions. This pattern is consistent with research showing that surface realism and anatomical plausibility are central to evaluations of digital human authenticity (Bae et al., 2025; Fraser et al., 2024; MacDorman et al., 2009; McDonnell et al., 2012), and that sensory facial features, particularly the eyes and skin, play a large role in anthropomorphic perception and mind attribution (Looser & Wheatley, 2010; Mazerant, van Berlo, et al., 2025; Waytz, Gray, et al., 2010). In line with cue-based accounts of anthropomorphism (Epley et al., 2007; Waytz, Cacioppo, & Epley, 2010), appearance seemed to function as an initial perceptual filter through which avatars were categorised as more human-like or as more artificial.

Participants also moved beyond static realism when evaluating avatars' capacity for emotional and social engagement. Despite both frameworks being animated with identical motion sequences, subtle eye movements and blinks were interpreted differently depending on which set of avatars was used. This suggests that perceived capacity for social and emotional engagement depended not only on animation itself but on how dynamic cues were implemented within each character rigging system underlying the mesh deformations. Prior work demonstrates that natural eye behaviour and coordinated micro-movements are central to impressions of vitality (Fu et al., 2008; Ruhland et al., 2015), and that disruptions to expressive dynamics can amplify uncanny impressions even when realism is high (Tinwell et al., 2011). The present findings extend this line of research by showing that identical motion can be experienced as natural or artificial depending on multiple factors such as surface detail and structural coherence. In this sense, mind attribution appears to depend on specific visual and dynamic cues, and is shaped by how observers interpret such combination of facial features and movement (H. M. Gray et al., 2007; Waytz, Cacioppo, & Epley, 2010).

Another theme concerned a more intangible sense of 'depth' or internal coherence. Participants frequently described avatars as feeling hollow, present, or unsettling without being able

to articulate what caused those impressions. Such descriptions align with perceptual mismatch accounts of the uncanny valley, which propose that discomfort arises when realism cues fail to align coherently (Di Natale et al., 2023; Kätsyri et al., 2015; MacDorman et al., 2009). In some cases, highly detailed surface rendering paired with an impression of a lacking emotional presence was experienced as eerie, suggesting that uncanniness can arise when outward human likeness does not support observers' inferences about the 'inner life' of a character. Simultaneously, these accounts align with mind perception theory, which distinguishes between perceiving an entity as human-like and perceiving it as having mental states (K. Gray & Wegner, 2012). Thus, participants' references to this intangible sense of 'depth' may reflect an emergent judgement formation about whether surface cues, structural plausibility, and implied psychological presence formed a coherent whole.

Overall, the qualitative findings suggest that impressions of both, humanness and uncanniness, may arise from the integration of surface realism, dynamic expressivity, and perceived internal coherence. Graphical fidelity influenced these impressions not simply by increasing realism, but by influencing how all the other human-like cues aligned to support inferences about mind and social presence of the digital human avatars.

3.2.5.3 Implications for the Uncanny Valley

The present findings, along with the results of *Study 1a*, clarify how graphical fidelity may relate to uncanny valley responses by showing that realism, mind attribution, and affective evaluation are related but distinct components of digital human perception. While higher fidelity MetaHuman avatars were rated as more humanlike and received stronger socio-cognitive attributions, the qualitative data revealed that these effects are graded rather than absolute. Realism appears to increase the likelihood of attributing emotional capacity and moral standing, but did not eliminate variability in how avatars were perceived overall.

This variability refines mismatch accounts of the uncanny valley, which propose that discomfort arises when realism cues fail to align coherently (Di Natale et al., 2023; Kätsyri et al., 2015; MacDorman et al., 2009). Participants' descriptions of avatars as visually detailed yet hollow illustrate how such mismatch may be experienced subjectively, consistent with mind perception theory distinguishing between perceiving an entity as humanlike and attributing mental states to it (Gray & Wegner, 2012).

The findings also provide additional nuance to the discourse about whether mind attribution explains uncanny responses. MacDorman (2024) argues that eeriness is often driven more strongly by more basic perceptual factors than by explicit mind attribution. The present results are compatible with this view: perceptual coherence appears to shape initial impressions, while socio-cognitive judgements form a related but partially independent evaluative layer. Further-

more, Higgins et al. (2023) reported that visual realism (LOD0 vs LOD5) did not significantly affect empathy or uncanny valley responses, despite earlier work with the same character models showing realism-related differences; they suggested this discrepancy may result from experimental design differences. More specifically, that realism was manipulated between subjects rather than within subjects and noted that expectation effects may be influential in within subjects uncanny valley measurement. The graded nature of the present effects similarly suggests that a repeated measures design may be better suited to detecting such differences, as direct comparison between avatar frameworks (or levels of detail in the case of Higgins and colleagues) likely heightens sensitivity to subtle variations in cue configuration that might be less apparent in between-subject designs due to complex individual differences in avatar perception and interpretation. Thus, their proposal that methodological factors may partly explain the mixed pattern of results appears reasonable in light of the present findings, as within participant designs can offer greater statistical sensitivity to graded effects (Charness et al., 2012; Montoya, 2023).

Taken together, the pattern of results observed across *Studies 1a* and *1b* indicates that advances in graphical fidelity between GFG and MetaHuman avatar frameworks do not function solely by lowering eeriness scores; rather, as evidenced by higher humanlikeness ratings for MetaHumans in *Study 1a* and their stronger attributions of emotional capacity, person (vs object) likeness, and moral concern in *Study 1b*, increased realism appears to shift the overall perceptual configuration towards a more humanlike interpretation. Although the quantitative effects were consistent, the qualitative findings indicate that participants' subjective impressions still varied in how those differences were experienced and interpreted. Thus, graphical fidelity seems to increase the likelihood of mind attribution without guaranteeing that all observers interpret the avatar in the same way.

3.2.6 Conclusion

This study demonstrates that graphical fidelity influences not only visual realism but also socio-cognitive interpretation. Higher graphical fidelity MetaHuman avatars were perceived as more humanlike and were attributed greater emotional capacity and moral concern than their lower graphical fidelity GFG counterparts, indicating that perceptual realism can shift the configuration of cues supporting mind attribution and other socio-cognitive judgements. Interestingly, these effects were graded rather than absolute. Qualitative findings revealed that impressions of humanness and uncanniness depended on the integration of surface realism, structural plausibility, and subtle dynamic expressivity. Realism alone did not appear to determine anthropomorphic impressions; coherence between the different human-likeness cues did.

By integrating quantitative and qualitative evidence, the present work shows that graphical fidelity shapes the conditions under which mind attribution becomes more likely, rather than

directly dictating how avatars will be perceived. Consequently, it provides a conceptual bridge between appearance-based uncanny valley research and the modelling of subtle behavioural cues that shape the perception of digital humans as psychologically present and socially meaningful.

The findings also point towards the need for more systematic investigation of how specific perceptual cues interact. Future work could experimentally manipulate particular facial regions, especially the eye region, to test their causal role in experience attribution and anthropomorphic perception. Subtle adjustments to eye texture quality, sclera visibility, gaze direction/behaviour, or structural integration within the surrounding facial features may help specify which cues most strongly contribute to impressions of psychological presence and inner life.

Participant gender was not examined as a potential moderator of pronoun attribution. Previous work suggests that observer gender can influence the perception of gender-related cues in virtual humans (Zibrek et al., 2015), and future research should examine whether male and female observers differ in the consistency with which they assign gendered pronouns to digital human avatars.

More broadly, the results suggest that dynamic expressivity should be isolated and examined independently from static realism. In the present study, motion sequences were held constant, yet they were interpreted differently depending on the avatar framework. This implies that dynamic cues do not operate in isolation, rather, they interact with surface and structural properties. The final empirical chapter of this thesis therefore holds graphical realism constant while systematically varying face motion, comparing stillness, highly unnatural motion, and highly natural motion, on measures of humanness, eeriness, and related uncanny valley dimensions. This design directly examines whether subtle variations in expressivity, independent of appearance, are sufficient to shift observers' evaluations of digital human characters.

Most importantly, the present findings motivate the reverse correlation approach administered in the following chapter. If mind attribution depends on coherent cue integration rather than realism alone, then modelling the precise temporal and parametric structure of facial motion becomes critical. The next chapter thus moves beyond comparing avatar frameworks and develops data-driven animation models of natural idle facial behaviour. By systematically manipulating facial dynamics obtained through reverse correlation, the thesis advances from evaluating realism to experimentally modelling the perceptual conditions under which digital faces appear 'alive'.

Chapter 4

Reverse Correlation of Natural Idle Facial Motion (Study 2)

4.0.1 Chapter Overview

This chapter uses a reverse correlation approach to model how observers perceive the naturalness of subtle, non-expressive idle facial motion in high-fidelity MetaHuman characters. Participants rated a large set of short, procedurally generated idle animations in which facial action units, head motion, eye saccades, and blinks were systematically varied. By relating these randomised motion parameters to naturalness judgements, the chapter derives data-driven generative models of low, medium, and high naturalness idle facial behaviour and tests their perceptual validity in an independent validation phase. The results show that naïve observers reliably distinguish animations synthesised from the low and high (and to some extent medium) naturalness models, supporting the claim that idle facial motion contains structured information that shapes impressions of aliveness and realism. Finally, the chapter outlines how these validated motion profiles enable motion naturalness to be manipulated while holding appearance constant, providing a methodological foundation for the next chapter’s test of motion-based uncanny valley responses.

4.1 Introduction

Creating lifelike digital human characters is a central goal of computer graphics and interactive media. As artificial agents become more humanlike in appearance and behaviour, they can also become susceptible to the uncanny valley effects, where greater realism sometimes leads to increased unease or discomfort in observers (Kätsyri et al., 2015; Mori, 1970; Mori et al., 2012). Such reactions can impact the believability of virtual humans by weakening the illusion of life

that is important for user engagement. Beyond the uncanny valley, understanding how subtle background movements contribute to the perceived aliveness of a character is also a general challenge for animation, since even neutral or idle states occupy a large proportion of a character’s on-screen time. A key challenge is determining what makes a digital human seem “natural” rather than uncanny. Recent work suggests that motion, including whether a character moves at all during “idle” moments, plays a decisive role in shaping how such agents are perceived and interacted with (Buisine et al., 2014; de Melo et al., 2010; Groom et al., 2009; Kocoń, 2013; Krämer et al., 2018). Studies show that subtle idle behaviours such as small postural shifts or looking around contribute to the believability and perceived humanlikeness of virtual agents and can modulate how deliberate motion conveying social signals, such as facial expressions, is perceived (Landa et al., 2025; Treall et al., 2020, 2021).

Within perceptual psychology, reverse correlation techniques have been developed to probe how people interpret and mentally represent complex stimuli such as faces and facial expressions. Reverse correlation is a data-driven method in which observers view randomly varied stimuli and their judgements are used to infer which features drive perception (see Jack & Schyns, 2017 for review). Reverse correlation provided important and unique insights into social cognition, revealing which facial features and movements people associate with personality traits and emotional expressions (Jack, Garrod, et al., 2012; Schyns et al., 2009; Todorov et al., 2011). To date, however, most reverse correlation studies have focused on overtly emotional or expressive stimuli and on personality-related judgements (Brinkman et al., 2017; Jack & Schyns, 2017; Todorov et al., 2025). There is a clear gap in applying reverse correlation to more subtle dynamics such as neutral or idle behaviours which lack strong expressions but are crucial for developing believable artificial human agents. Idle motion, the minute, non-communicative movements present even when one’s face or body is “at rest” or neutral, may be particularly important for making virtual humans appear alive (Kocoń, 2013; Treall et al., 2020, 2021); additionally, it has been highlighted as a significantly understudied area in perceptual research more broadly (Landa et al., 2025). Landa and colleagues also showed that while acted and “genuine” idle motions are perceptually indistinguishable, handmade animations can be reliably told apart from motion-captured ones, emphasising both the scarcity of high-quality idle motion capture datasets and the challenges of crafting believable idle behaviour manually. In this context, deriving generative models of idle motion directly from observers’ perceptual judgements, using reverse correlation paradigms, offers a complementary route to synthesising natural idle facial motion without relying exclusively on large motion-capture datasets or extensive manual labour.

In the following sections we integrate research across these themes, laying the groundwork for the present study, which uses reverse correlation to identify the facial action units and motion parameters that make an idle animated face look natural. Although work on neutral and idle behaviour remains relatively sparse compared with the vast literature on expressive faces, this

reflects the fact that idle motion has only recently been treated as a distinct and theoretically important target for perceptual research (Landa et al., 2025). First, we revisit key developments in uncanny valley and adjacent research, focusing on how motion, and in particular the presence or absence of idle behaviour, shapes responses to digital humans.

Next, we examine the use of reverse correlation in perceptual and social cognition research, highlighting how people infer emotions and traits from facial expressions and movement. We then discuss how reverse correlation work has predominantly investigated overtly emotionally expressive stimuli, and why neutral idle motion remains an underexplored but critical frontier for believability in artificial human agents. Finally, we outline the reverse correlation paradigm adopted in the present study, in which participants rate large sets of procedurally generated idle facial animations, allowing us to pinpoint which patterns of facial movement are associated with higher perceived naturalness. Beyond characterising “natural” idle behaviour, the resulting data-driven motion model provides a parametric space in which idle motion can be systematically manipulated in the following chapter to test how specific changes in movement dynamics affect uncanny valley-related measures. We conclude by situating this approach within broader efforts to synthesise realistic motion for digital humans and to extend reverse correlation techniques to more subtle, everyday aspects of human behaviour.

4.1.1 Uncanny Valley and the Role of Motion in Human-likeness

The uncanny valley, first proposed by Masahiro Mori (Mori, 1970; Mori et al., 2012), describes how increasing human likeness in artificial agents can lead to a sharp increase in discomfort when an entity becomes almost, but not entirely, human. Contemporary accounts suggest that such reactions are driven less by human likeness specifically than by perceptual mismatches between cues that are typically coherent in real human perception (Kätsyri et al., 2015). Inconsistency between appearance and behaviour is hypothesised to produce a perceptual mismatch, for example, a highly realistic face paired with stiff or unnatural movement. This mismatch perspective is especially relevant for the present chapter as it suggests that uncanniness is driven not only by how human a character looks, but by whether its movement matches observers’ expectations of living humans. Thus, the focus here is on the dynamic features of human likeness: whether motion is perceived as natural and realistic rather than unnatural or artificial.

4.1.1.1 Motion as a Modulator of Uncanny Perception

Motion that aligns with perceptual expectations for natural movement, characterised by smooth and coordinated dynamics, may enhance perceived realism and human likeness. Likewise, motion that is awkward, overly rigid, or temporally misaligned may elicit strong uncanniness.

Experimental manipulations introducing subtle temporal asynchronies into facial animations demonstrate that even small disruptions to the timing and coordination of movement can reliably increase perceived uncanniness, and that these effects occur across human videos, robotic faces, and computer-generated avatars (Diel et al., 2023). These findings align with mismatch-based accounts of the uncanny valley and are consistent with predictive processing models that emphasise violations of expected human motion, which propose that observers continuously generate expectations about how human faces and bodies should move, and experience discomfort when those expectations are violated (Diel & Lewis, 2022).

Importantly, motion quality extends beyond obvious, overt expressions to the fundamental way in which movement unfolds over time, such as its smoothness, rhythm, and natural variability. These dynamic features of movement contribute towards whether motion is perceived as humanlike or artificial. Motion that looks stiff, or mechanical can undermine the impression of agency and aliveness, whereas biologically plausible dynamics tend to increase affinity and engagement. For instance, characters animated with natural human movement are typically judged as more acceptable and likeable than those animated with movement that has been altered to look more artificial (Piwek et al., 2014); Notably, Williams et al. (2019) found that natural motion attracted greater attention regardless of whether the agent was human or non-human, suggesting that motion itself is an important cue to animacy. Thus, indicating that perceptual expectations about movement may also operate independently of appearance, and that violations of these expectations may contribute to uncanniness even in the absence of a very humanlike form. Together, these findings suggest that motion may contribute to uncanny responses not through its presence alone, but through the extent to which its dynamics conform to expectations for natural human movement.

4.1.1.2 Facial and Bodily Motion in the Uncanny Valley

Much of the uncanny valley literature traditionally focused on facial appearance and expression, given the central role of faces in identity, emotion, and social communication (Schindler et al., 2017). However, motion of the body and head also plays a critical role in whether an artificial agent is perceived as alive. Even when facial appearance is highly realistic, a completely still body can be unsettling. Observers implicitly expect living beings to exhibit continuous, low-level movement such as subtle postural sway or small adjustments that reflect ongoing bodily regulation. When these cues are absent, characters may come across as artificial.

Empirical evidence supports the intuition that subtle bodily motion contributes to the perceived aliveness of virtual humans. Treval et al. (2021) demonstrated that adding natural postural oscillations to an avatar's torso, derived from human motion capture and intended to simulate the small sway associated with balance and breathing, enhanced both the perceived intensity and

believability of a facial pain expression. Participants judged the avatar as more convincing and reported stronger affective responses when this background bodily motion was present, compared to when the body remained completely static. Treval and colleagues further argued that pairing a dynamically emoting face with a completely still body may introduce a cross-channel inconsistency that could contribute to eeriness in virtual agents; however, this suggestion is presented as a theoretical rationale rather than an empirically tested outcome in their study, which focused on pain perception and affective response rather than uncanny impressions.

This effect has some analogues in everyday experience. For example, a street performer posing as a “living statue” often captures attention because they violate expectations for continuous human movement. Observers may find themselves waiting for a blink or a slight shift as confirmation that the figure is alive. In these ambiguous situations, subtle movements that normally go unnoticed become the decisive indicators of aliveness. The same expectations come to play when someone is asleep: even when they are otherwise still, we still anticipate small signs of ongoing bodily regulation such as breathing or subtle posture shifts, and their absence can feel worryingly unnatural.

Findings from robotics research further emphasise the importance of idle behaviour. Asselborn et al. (2017) demonstrated that a humanoid robot exhibiting subtle idle movements during moments of inactivity was perceived as more humanlike and friendly than an otherwise identical robot that remained static during moments of inactivity. These results suggest that ongoing, low-level movement functions as a basic cue to aliveness, rather than as an explicit social signal. By contrast, the absence of idle motion led the robot to be described by the participants as “*like a statue... frozen while waiting*”. These findings also align closely with long-standing animation practice, in which idle behaviours such as blinking or subtle posture shifts are regularly added to prevent characters from appearing inert. More broadly, Bates (1994) describes believability as the “*illusion of life*” achieved when behaviour makes a character feel like it has an ongoing internal state.

Taken together, this body of work highlights that motion is a double-edged sword in the uncanny valley. Natural, humanlike movement can noticeably increase an artificial character’s perceived humanlikeness and appeal, whereas more artificial motion, or complete absence of expected movement, could make an otherwise acceptable character come across as uncanny. Across studies, several consistent themes emerge: the importance of temporal coherence in movement, the role of biologically plausible dynamics, and the contribution of continuous, low-level idle behaviour to perceptions of animacy and humanlikeness. These insights suggest that addressing the uncanny valley problem is not simply a matter of improving how characters look, but also of understanding how certain patterns of motion support the impression of a living agent. In the following section, we turn to a methodological approach well suited to probing these problems directly; more specifically, the method of reverse correlation can use observers’

perceptual judgements to identify which features of idle facial motion systematically contribute to perceived naturalness of motion.

4.1.2 Reverse Correlation in Social Perception Research

4.1.2.1 Overview of Reverse Correlation

Reverse correlation is a psychophysical method used to discover which stimulus features influence human judgments, without requiring strong a priori assumptions about those features. Psychophysics broadly refers to approaches that systematically relate controlled variations in physical stimulus properties to changes in subjective perception, allowing perceptual sensitivities to be inferred from behavioural responses (Ahumada & Lovell, 1971; Jack & Schyns, 2017). In social face perception, reverse correlation is typically implemented by (1) selectively revealing information already present in a stimulus, (2) adding random noise to drive emergent perceptions, or (3) sampling from higher-order generative feature spaces, such as facial action units and their dynamics, to model the information that is driving the judgments.

In an early demonstration, Gosselin and Schyns (2003) applied reverse correlation to pure noise stimuli to probe the perception of a “happy face”. Participants were shown random visual noise patterns and asked to report whether they perceived a smiling face. Importantly, while the stimuli contained a faint outline of a face, no smiling mouth was ever presented in any of the images. By averaging the noise from trials in which participants reported perceiving a happy face, the researchers revealed a coherent smiling facial shape. In other words, human perception can extract a meaningful face signal from static noise. This finding suggests that observers carry internal templates of facial expressions, and that reverse correlation can externalise these underlying cognitive representations without presenting the target signal itself.

Beyond basic emotions, reverse correlation has been widely used to investigate how people infer more complex social traits from faces, one such example being perception of trustworthiness. Humans form rapid impressions of whether a face appears trustworthy or dominant, and these impressions are thought to arise from subtle resemblances to emotional expressions; for instance, faces resembling happiness tend to be judged as more trustworthy, whereas faces resembling anger are judged as more dominant or untrustworthy (Oosterhof & Todorov, 2009; Todorov, 2008). Using a reverse correlation paradigm, Dotsch and Todorov (2012) asked participants to repeatedly choose which of two randomly generated faces appeared more trustworthy. Averaging across selected faces produced classification images that depicted the visual patterns underlying these judgments, demonstrating how reverse correlation can reconstruct observers’ mental representations of social traits.

Building on this approach, reverse correlation has since been used to explore a wide range of higher-level social inferences, including stereotypes and prejudice, dominance, extraversion, and even political ideology, revealing the facial configurations that drive these social impressions (Brown-Iannuzzi et al., 2017, 2018; Oliveira et al., 2019; Young et al., 2014). Across this body of work, reverse correlation methodologies show that many social judgements rely on consistent patterns of facial information salient to observers, making it possible to visualise how beliefs, attitudes, and trait impressions are reflected in perceived facial appearance.

4.1.2.2 Reverse Correlation with Dynamic Stimuli

While many early reverse correlation studies relied on static images, more recent work has extended the approach to dynamic facial motion (Jack & Schyns, 2017). One way to do this is to generate animations by randomly varying the activation of different facial action units (FACS; Ekman and Friesen, 1978) over time, then use observers' responses to determine which patterns of facial movement drive particular perceptions. Yu et al. (2012) introduced this approach whereby they reverse correlated such dynamic face stimuli. In their study, they created a large set of short 3D facial animations, each animation comprised of random combinations of FACS-based facial movements, with each individual movement controlled by a parameterised activation curve. Observers categorised each animation as one of the six basic emotion categories (happiness, surprise, fear, sadness, anger, and disgust; Ekman et al., 1972) or as "other", and also rated its emotional intensity. By relating the random action unit activations to emotion labels, and the temporal parameters of the activation curves to intensity judgements, Yu and colleagues obtained perception-driven models of dynamic facial expressions. This work shows that reverse correlation can be applied to dynamic stimuli changing over time to recover not only which facial movements matter for emotional expression perception, but also how their timing and intensity shape what observers perceive.

This approach to reverse correlation with dynamic facial stimuli has been applied to a range of facial social perception contexts. As reviewed by Jack and Schyns (2017), generative approaches to facial motion have been used to model not only emotions (Jack, Garrod, et al., 2012; Jack et al., 2014, 2016), but also other mental states such as confusion, interest, and boredom, as well as trait impressions such trustworthiness, dominance, and attractiveness (Chen et al., 2015; Gill et al., 2014). The method was also applied to investigate cultural differences in emotion perception (Jack, Garrod, et al., 2012). A key advantage of these approaches is that they allow facial movement patterns to be revealed directly from observer responses, rather than being constrained by predefined expression templates or theoretical categories.

However, despite this increased flexibility, most reverse correlation studies in face perception, both static and dynamic, have focused on relatively overt, interpretable states of mind or

social meaning (Jack & Schyns, 2017; Todorov et al., 2025). Whether targeting emotional expressions, personality traits, or communicative mental states, these paradigms typically involve facial movements that are salient, expressive, and clearly linked to specific social signals. By contrast, as far as we are aware, at the time of writing no reverse correlation work has been carried out on the perceptual role of subtle, non-expressive facial movements that occur when a face is neutral or at rest. Such idle facial motion does not convey explicit emotion or intention yet may play a critical role in shaping impressions of aliveness, naturalness, and realism. The relative absence of this kind of movement from reverse correlation research highlights an important gap in our understanding of how dynamic facial information contributes to face perception.

4.1.3 Reverse Correlation, Idle Motion, and the Present Study

Taken together, the reviewed literature suggests that subtle idle movement functions as a foundational cue to aliveness and natural behaviour in human perception. However, despite its recognised importance in animation practice, robotics, and perceptual studies of motion, reverse correlation research has yet to target idle facial motion directly. Existing work has largely focused on overt expressions, communicative mental states, or trait impressions (Jack & Schyns, 2017; Todorov et al., 2025), leaving the perceptual role of neutral facial dynamics largely unexplored. At the same time, studies of idle behaviour show that background movement can influence how social signals are perceived (Treal et al., 2020, 2021), and recent work highlights idle motion as both perceptually impactful and understudied, with practical challenges in obtaining high-quality motion data or crafting believable animations manually (Landa et al., 2025). Together, these observations motivate a data-driven approach: applying reverse correlation to idle facial motion allows perceptually relevant movement patterns to be derived directly from observer judgements and used to synthesise natural-looking idle behaviour.

Thus, the present study applies reverse correlation to identify which patterns of subtle, non-expressive idle facial motion drive perceptions of naturalness. Rather than focusing on overt emotional expressions or communicative signals, the study targets the low-level facial dynamics that occur when a face is neutral yet perceptually alive. Participants are presented with large sets of procedurally generated idle facial animations in which facial action units and their temporal parameters are systematically varied, and are asked to rate the naturalness of each animation. By relating these perceptual judgements to the underlying motion parameters, the study derives data-driven models of idle facial behaviour that reflect the movement patterns observers associate with natural-looking neutral faces.

Importantly, this approach allows idle motion to be examined without relying on predefined animation rules or large motion capture datasets. Instead, naturalness is defined through observers' perceptual responses, enabling the reconstruction of idle facial motion directly from

human judgements. In doing so, the study extends reverse correlation methodology beyond expressive facial behaviour to a more subtle, yet potentially perceptually critical, aspect of facial motion.

4.2 Methods

4.2.1 Participants

All participants were recruited through the University of Glasgow’s participant pool platform. The participation criteria included being aged 18 or older, having normal or corrected-to-normal vision, being free of any: 1) lexical, reading, language, hearing atypicality (e.g., dyslexia); 2) social/emotion processing related atypicality (e.g., autism spectrum disorder, depression, anxiety); 3) any face processing atypicality (e.g., prosopagnosia). In total, 28 participants were recruited for the two phases of this study. The first phase (reverse correlation) was conducted with 18 participants (7 male, 11 female) with a mean age of 25 ($SD = 4.96$). The second phase (validation) was conducted with a separate set of 10 participants (3 male, 7 female), with an average age of 31.4 ($SD = 8.09$). All participants met the participation criteria. Each participant received a £10 Amazon Voucher as compensation for taking part in the study. The study received ethical approval from the University of Glasgow’s College of Medical, Veterinary & Life Sciences Ethics Committee (Ethics Code: 200240259).

4.2.2 Materials

Python (v3.7; Van Rossum & Drake Jr, 1995) was used for generating animation curve data, collecting participant responses, and stimuli presentation along with the PsychoPy library (Peirce et al., 2019). Animation and video rendering were conducted in Unreal Engine 5.6 (Games, 2025). Video rendering was also aided by the FFmpeg library (Tomar, 2006). The animation rendering pipeline was ran on a Windows 11 desktop equipped with an RTX3070 GPU, 32GB RAM, and a Ryzen 5 5600 CPU.

The experiment was conducted on an M1 MacBook Pro (2021) connected to a 24” full HD (1920x1080p) 60Hz monitor via a HDMI cable. The monitor was used as the main display with the MacBook placed out of participants’ sight. Participants provided their responses using a USB-wired keyboard. Additionally, a chin rest was used to maintain a consistent visual viewing angle across participants (approximately 8° ; see *Fig.4.1*).

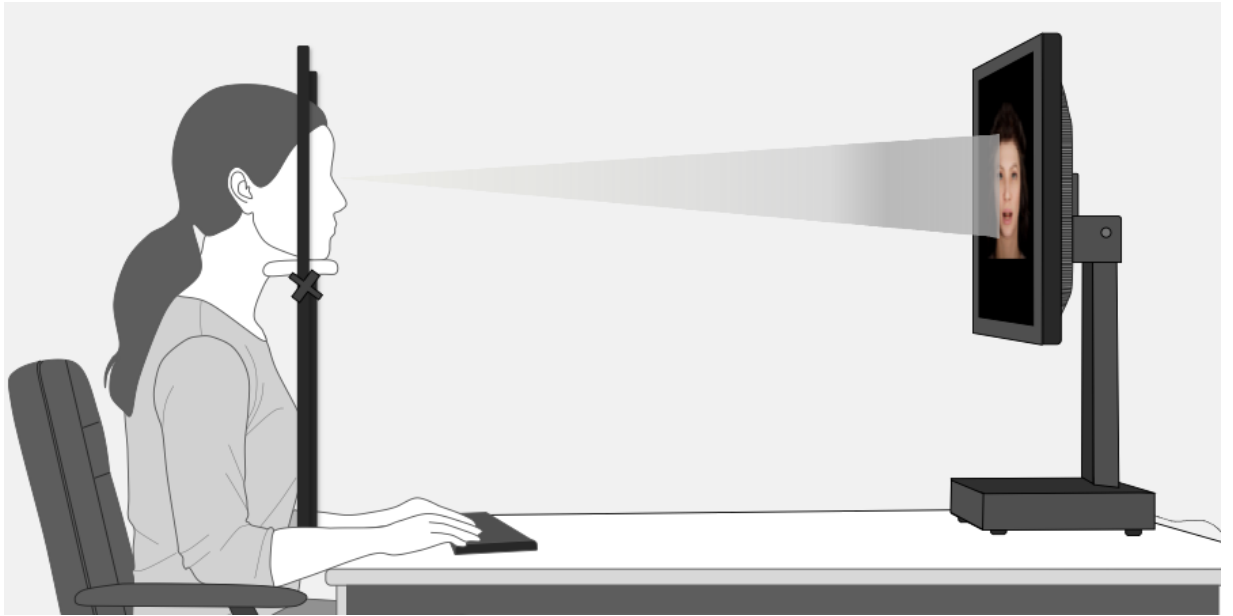


Figure 4.1: *Experimental setup used for stimulus presentation and response collection. Participants were seated at a fixed viewing distance using a chin rest to maintain a consistent visual angle (approximately 8°) while viewing stimuli on a 24" monitor. Responses were provided via a keyboard, with the laptop positioned out of sight to minimise distraction and ensure standardised viewing conditions across participants.*

4.2.3 Stimuli

The stimuli set consisted of videos depicting animations across 15 different MetaHuman Unreal Engine characters (Epic Games, 2021), with 8 male and 7 female avatars. The characters also varied in age and race to span a wide range of identities (see *Fig.4.2*).

In total, 1500 unique animations were produced for the stimulus set, with 100 animations generated for each of the 15 unique characters. All animations were 4 seconds-long and rendered at 60 frames-per-second (240 frames per animation). This set of animations extensively sampled the space of the dynamic parameters. Stimuli were generated using the custom Unreal Engine 5.6 animation framework described in *Chapter 2*.



Figure 4.2: *MetaHuman identities used to generate the stimulus set. The full set comprised 15 characters (8 male, 7 female) varying in age and ethnicity to capture a broad range of facial identities. All characters were created using Unreal Engine’s MetaHuman framework and served as the base identities for the animated stimuli. This diversity was included to ensure that the results were not driven by a specific face or demographic category.*

4.2.3.1 Action Units

To generate each animation, a small subset of facial action units (AUs) was selected at random from a pool of 45 AUs, resulting in a set of unique AUs for each animation. The number of AUs selected per animation was determined by drawing a value from a normal distribution ($M = 5$, $SD = 1$).

For each AU in the selected subset, an activation curve was generated, where a value of 0.0 corresponded to no activation and a value of 1.0 corresponded to maximal activation. The temporal activation curves were generated using a function derived from Yu et al. (2012), with the addition of a sustain parameter. This function was parameterised by seven values: peak amplitude, peak latency, onset latency, offset latency, acceleration, deceleration, and sustain (see Fig.4.3).

$$\text{Peak Amplitude} = \text{CP2}_y = \text{RV1}$$

$$\text{Peak Latency} = \text{CP2}_x = 0.2 + (0.67 - 0.2) * \text{RV2}$$

$$\text{Onset Latency} = \text{CP1}_x = \text{CP2}_x * 0.67 * \text{RV3}$$

$$\text{Offset Latency} = \text{CP3}_x = (\text{CP2}_x * (1 - 0.67) + 0.67) * (1 - \text{RV4}) + \text{RV4}$$

$$\text{Acceleration} = \text{CP1}_y = \text{CP1}_x * \text{CP2}_x * \text{RV5}$$

$$\text{Deceleration} = \text{CP3}_y = (1 - \text{CP3}_x) * \text{CP2}_y * \text{RV6}$$

Figure 4.3: *Parameterisation of facial action unit (AU) activation curves used for animation synthesis. The temporal profile of each AU was defined by seven parameters: peak amplitude, peak latency, onset latency, offset latency, acceleration, deceleration, and sustain. These parameters were generated from a set of random variables (RV1-RV6) and combined using a function adapted from Yu et al. (2012), with the addition of a sustain component. The sustain parameter controls the duration for which the AU remains at its peak activation before returning to the baseline. This formulation defines the timing and shape of each AU activation curve, which was then used to drive facial motion in the generated animations.*

Each of these parameters was sampled independently from a uniform distribution ranging between 0.01 and 1.0, with the exception of peak amplitude, which was sampled from a reduced uniform range between 0.01 and 0.5 to constrain overall movement magnitude of an idle face. The sustain parameter controlled the duration for which the activation curve remained at peak amplitude before transitioning into the offset phase, extending the original formulation by Yu et al. (2012) while preserving the overall temporal structure of the activation function.

Following Yu et al. (2012), each activation curve was constructed using a cubic Hermite

spline defined by a small set of control points that specify how the AU activation changes over time. The six random values sampled from uniform distributions were transformed into parameters that determined when the movement begins, when it reaches its peak, and when it returns to zero. These transformations ensured that the resulting curves followed a reasonable temporal order (with onset occurring before peak, and peak before offset) and produced smooth motion. The additional sustain parameter introduced here simply inserts a short holding period at peak activation before the curve begins to offset, allowing the model to produce brief sustain of activation. *Figure 4.4* provides example curves illustrating how these parameters may shape the resulting activation profiles.

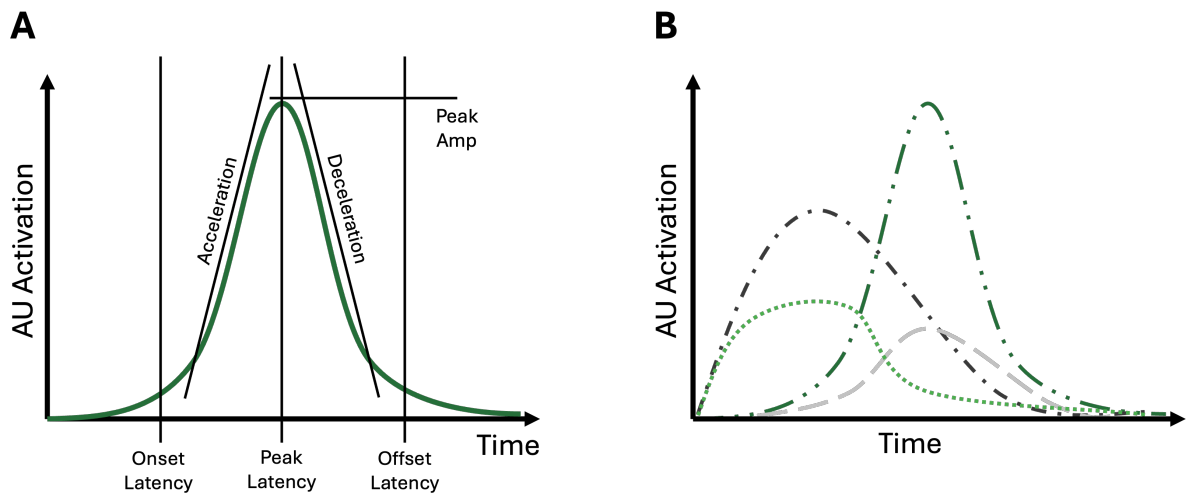


Figure 4.4: *Illustration of action unit (AU) activation curve parameterisation and example curves. (A) Schematic of a single AU activation curve showing how the control point parameters define its temporal structure, including onset latency, peak latency, offset latency, peak amplitude, and the acceleration and deceleration phases. (B) Example activation curves demonstrating how variation in parameter values may produce different temporal profiles of AU activation.*

4.2.3.2 Head Motion

In addition to facial action unit activity, each animation included procedurally generated idle head motion. Head motion was modelled as slow, continuous rotational movement around the X (pitch), Y (yaw), and Z (roll) axes. For each axis, an independent motion curve was generated using a custom stochastic curve generation function designed to approximate naturalistic idle head behaviour.

For each animation, head motion along each axis was controlled by two parameters: movement magnitude and smoothness. Magnitude determined the maximum rotational deviation (in degrees) from the neutral head position, while smoothness controlled the temporal smoothing applied to the motion curve. Motion curves were generated by combining periods of relative

stability with occasional directional changes, producing movement with occasional pauses and directional changes. Temporal smoothing was applied using a Gaussian filter to remove abrupt frame-to-frame transitions and ensure biologically plausible motion trajectories.

To avoid abrupt onsets and offsets, all head motion curves were tapered at the beginning and end of the animation, such that the head returned to a neutral, forward-facing position. Head motion spanned the full 4-second duration of each animation and was generated such that all motion curves began and ended at a neutral, forward-facing head position, ensuring that all animations started and finished in the same pose.

The magnitude and smoothness parameters for head motion were randomly sampled on each trial and included as part of the motion parameter set entered into the reverse correlation analysis. The magnitude parameter was sampled in the following ranges for: X-axis (0.1 to 1.5), Y-axis (0.1 to 1.5), Z-axis (0.1 to 1.0). The smoothness parameter was sampled in the following range of 0 to 20 for all three axis. This allowed the contribution of different head motion characteristics to perceived naturalness to be modelled alongside facial action unit dynamics.

4.2.3.3 Eye Saccades

Each animation also included procedurally generated eye saccades to simulate gaze behaviour. Saccades were modelled as rapid shifts in gaze direction along the horizontal (X) and vertical (Y) axes. Saccade generation was controlled by two parameters: frequency and magnitude. Frequency determined the number of saccadic events per second, while magnitude controlled the maximum deviation of each eye movement from the neutral position. The frequency parameter was sampled from the following list of values for both axes: 0.5, 0.75, 1.0, 1.25, 1.5. The magnitude parameter was sampled in the following range of 0.005 to 0.3 for both axes. For each animation, a fixed number of saccades was generated based on the selected frequency parameter, and saccade onset frames were distributed randomly across the animation duration.

For each saccade event, a target gaze direction was sampled uniformly and scaled by the selected magnitude parameter. Between saccades, gaze position was held constant, and Gaussian smoothing was applied to the gaze trajectory to produce rapid yet continuous transitions between fixation points. This resulted in brief gaze shifts separated by periods of relative stability, consistent with natural eye movement dynamics during rest.

Saccade motion was generated for the full 4-second duration of each animation and did not repeat. As with head motion, all saccade trajectories began and ended at a neutral forward gaze position, ensuring consistent initial and final eye positions across all stimuli.

Saccade frequency and magnitude parameters were varied randomly across animations and included in the reverse correlation analysis. This enabled assessment of how different patterns

of idle gaze behaviour contributed to observers' perceptions of facial motion naturalness.

4.2.3.4 Blinks

Lastly, each animation also included procedurally generated eye blinks to provide a further cue to aliveness during idle states. Blinks were modelled as brief, bilateral eyelid closures. Blink generation was controlled by two parameters: frequency and smoothness. Frequency determined the number of blinks occurring over the animation duration, while smoothness controlled the temporal sharpness of eyelid opening and closing. Higher smoothness values produced slower, more gradual blinks, whereas lower values resulted in faster, more snappy motion.

For each animation, a fixed number of blink events was generated based on the selected frequency parameter, and blink onset frames were distributed randomly across the animation duration. Blink amplitudes were scaled to ensure full (or almost full) eyelid closure, and Gaussian smoothing was applied to the blink time series to produce continuous, biologically plausible eyelid motion. Blink frequency and smoothness parameters were varied across animations and included in the reverse correlation analysis. The frequency parameter was sampled from the following list of values: 0.25, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0. The smoothness parameter was sampled in the following range of 4.0 to 10.0.

Blink motion was generated for the full 4-second duration of each animation and did not repeat. All blink sequences began and ended with the eyes fully open, ensuring that all animations started and finished in the same neutral eye state. See *Figure 4.5* for a visualisation of the temporal structure of each stimulus.

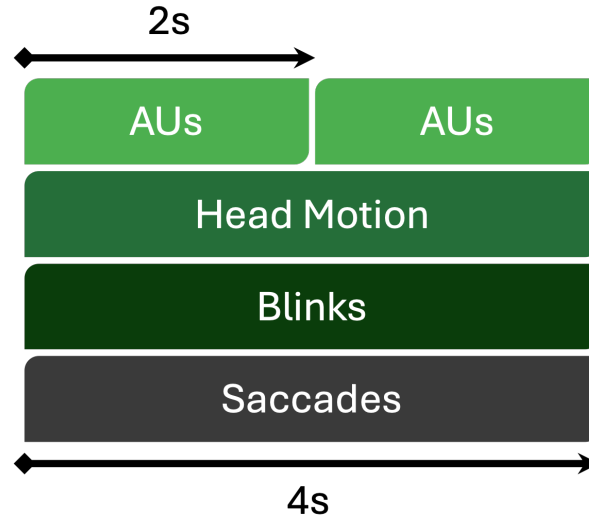


Figure 4.5: *Temporal structure of the procedurally generated animation components in the final stimuli. Each stimulus consisted of multiple layers of motion generated over a 4-second duration. Head motion, eye saccades, and blinks were present continuously throughout the full 4-second animation, while facial action unit (AU) activity was embedded as a shorter 2-second segment that was presented twice within the sequence.*

4.2.3.5 Validation Stimuli

All validation animations were generated directly from the parameters obtained in the reverse correlation analysis. The validation stimulus set comprised of 90 animations generated across the 15 digital human characters used in the reverse correlation phase (*Fig.4.2*), with six animations produced per character: two corresponding to low naturalness, two to medium naturalness, and two to high naturalness, resulting in 30 animations per naturalness level in total. Each animation was 6 seconds long (360 frames at 60 fps) and contained continuous idle head motion, eye saccades, and blinks across the full duration of the clip. Facial action unit (AU) activity was embedded as a single 2-second facial animation segment occurring midway through the video (from 2 to 4 seconds), with no repetition (*Fig.4.6*). For each animation, a small subset of facial AUs was selected at random from the pool of action units identified as significant for the corresponding naturalness level in the reverse correlation analysis (see results section for more details). The number of AUs included in each animation was determined by sampling from a normal distribution ($M = 3$, $SD = 1$), resulting in a unique combination of AUs for each clip.

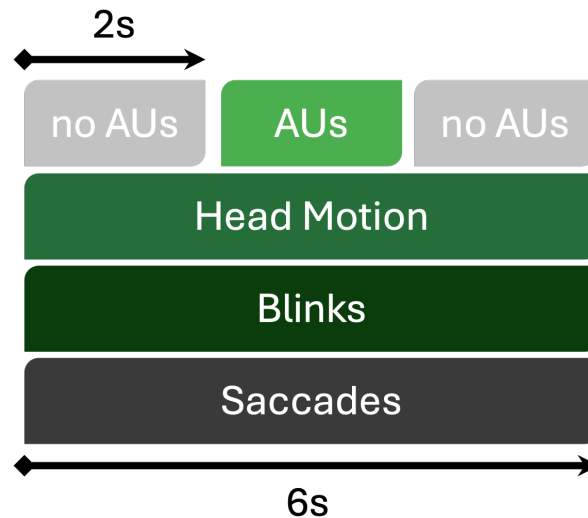


Figure 4.6: *Temporal structure of validation stimuli. Each validation animation lasted 6 seconds, with continuous head motion, eye saccades, and blinks present throughout the full duration. Facial action unit (AU) activity was embedded as a single 2-second segment occurring midway through the animation (2-4 seconds), with no repetition.*

4.2.4 Procedure

At the start of the study, the experiment instructions were provided on screen for the participant to read. The instructions were as follows: “In this experiment, you will watch short videos of digital human avatars. The avatars will be moving in a neutral/idle manner, as if they are simply present without expressing any specific emotions or messages. Your task is to rate how natural their idle/neutral movements appear to you using a scale from 1 (“*Very Unnatural*”) to 7 (“*Very Natural*”). It is important that you rate the motion and not the physical appearance of the characters. During the task try to use the full range of this scale when providing your ratings. Try to ensure that your decisions are based on all aspects of motion (head motion, eye motion, and face motion). If any of these affect your judgement, please reflect this in your ratings”.

Participants were intentionally asked to provide an overall judgement of how natural each animation appeared. Although perceived naturalness has previously been used to evaluate animation (e.g., Lei et al., 2021; Pollick et al., 2005; Voas et al., 2023), there is currently no widely accepted operational definition of perceived naturalness of motion. Previous work has similarly used naturalness as an intuitive perceptual judgement made by the observer rather than as a strictly defined objective property (Voas et al., 2023). Thus, participants were not provided with a formal definition of naturalness or placed within a specific scenario. Instead, the judgement prompt was intentionally kept simple and left open to their own interpretation.

This approach was adopted to maximise ecological validity, as people do not typically eval-

uate facial behaviour against explicit criteria during everyday social interactions. Instead, they form an immediate impression of whether another person's behaviour appears natural or unnatural. Providing additional context or a specific scenario may have encouraged participants to focus on particular aspects of the animation or adopt more constrained evaluation criteria, rather than allowing their spontaneous perception of motion naturalness to guide their responses. The present study therefore aimed to capture these intuitive judgements and use reverse correlation to identify the facial movement characteristics associated with perceived naturalness. Before commencing with practice trials, participants were verbally reminded to focus on the motion rather than the appearance of the characters when providing naturalness ratings, and were provided with an opportunity to have any question answered.

At the start of the experiment, participants completed 10 practice trials with the researcher present to familiarise themselves with the task and have any further questions answered. Upon completion of the practice trials, the researcher left the room and participants commenced with the experimental trials.

For each participant the 1500 trials were presented in a fully randomised order. In each trial, the 4-second video plays on a loop without a time limit until a naturalness rating is provided. Following submission of a rating is a 100ms intertrial interval with a white cross in the middle of a black background, followed by the start of the next trial's video presentation. Although no time limit for response was specified, the participants were verbally instructed to try and provide their responses based on first impressions whenever possible. After every 500 trials, the participants were provided with a break of up to 5 minutes. Upon completion of all 1500 trials participants were debriefed and the study concluded.

To validate the obtained models of natural idle motion from the reverse correlation portion of the experiment, the other 10 "naïve" participants were presented with 90 animations synthesised from the reverse correlation data. For each one of the 15 digital human characters, we generated 6 new animations: 2 animations based on naturalness rating of 1 (*"Very Unnatural"*), 2 animations based on rating of 4 (middle naturalness), and 2 animations based on rating of 7 (*"Very Natural"*), resulting in 30 animations per level of naturalness (low, medium, high). The procedure was exactly the same as with the reverse correlation phase, with the main difference being a break of up to 5 minutes every 30 trials. The intention was to see whether naïve observers would rate the videos congruently with the level of naturalness used to generate the different animations.

4.3 Results

4.3.1 Data Analysis

To identify the facial action units (AUs) most associated with perceived naturalness, we conducted a multi-step analysis. First, for each participant, we computed Pearson correlations between binary AU activations (whether the AU is present or not on a trial) and each naturalness rating (1-7), identifying AUs that showed significant positive correlations ($p < .05$) for each naturalness rating. Secondly, we aggregated these results across participants, counting how many participants showed significant correlations for each AU at each rating to see what AUs tend to be associated with what ratings across the group. If either the left or the right side of a given AU was significant (i.e., the left or right lateralised variant of that AU), the AU was treated as present bilaterally and marked as significantly associated with that rating. Thirdly, we grouped naturalness ratings into low (1-2), mid (3-5), and high (6-7) naturalness and calculated how many participants showed significant AU correlations within each of these naturalness categories. To ensure that only action units showing consistent associations across observers were included in the generative modelling stage, we retained AUs that were significantly associated with a given naturalness level in four or more participants (22% of the sample). This filtering step reduces the influence of AUs that appear significant in only a small number of participants, while still retaining a broad set of candidate AUs for animation synthesis. For all participants as a group, regression modelling was also used to model how motion parameters for head motion, eye blinks, and saccades change as a function of naturalness ratings (1-7).

To further test whether the prevalence of these AUs differed systematically across naturalness levels, we conducted Cochran's Q tests for each AU using participant-level binary indicators of whether a given AU was significantly associated with each naturalness category (low, mid, high). Where significant effects were detected, post hoc exact McNemar tests were used to determine which naturalness levels differed from each other. In simple terms, this analysis tested whether certain facial movements tended to appear significantly more often in one naturalness category than in others across participants.

4.3.2 Action Units

Some of the most consistently significant AUs for low naturalness (score 1-2) included Jaw_Y - AU26 (11/18 participants; 61%), Mouth_UpperLipRaise - AU10 (9/18 participants; 50%), Mouth_CornerPull - AU12 (9/18 participants; 50%) and Mouth_LowerLipDepress - AU16 (8/18 participants; 44%).

Mid naturalness (score 3-5) was associated with a wider range of moderately frequent AUs with most frequent including Upper_LipRaise - AU10 (7/18 participants; 39%), MouthFunnelUpper - AU22 (7/18 participants; 39%) and MouthFunnelLower - AU22 (6/18; 33%).

High naturalness (score 6-7) also showed moderate levels of consistency, with some of the most frequently significant AUs including MouthLipPressUpper - AU24 (7/18 participants; 39%), MouthTightenUpper - AU23 (7/18 participants; 39%), and BrowRaiseOuter - AU2 (6/18; 33%). See *Figure 4.7* for more details on proportions of significant AUs.

Cochran's Q tests indicated that the prevalence of some AUs differed significantly across naturalness levels, including Jaw_Y (AU26; $Q(2) = 13.29$, $p = .001$), Mouth_UpperLipRaise (AU10; $Q(2) = 10.31$, $p < .01$), and Mouth_CornerPull (AU12; $Q(2) = 10.17$, $p < .01$). Post hoc exact McNemar tests showed that all three AUs were significantly less prevalent for high naturalness than for low naturalness (all $p < .01$). Mouth_UpperLipRaise was also significantly less prevalent in high compared to mid naturalness ($p < .05$), whereas other comparisons involving mid naturalness were not significant.

Frequency of Action Units Significantly Associated with Perceived Naturalness Across Participants

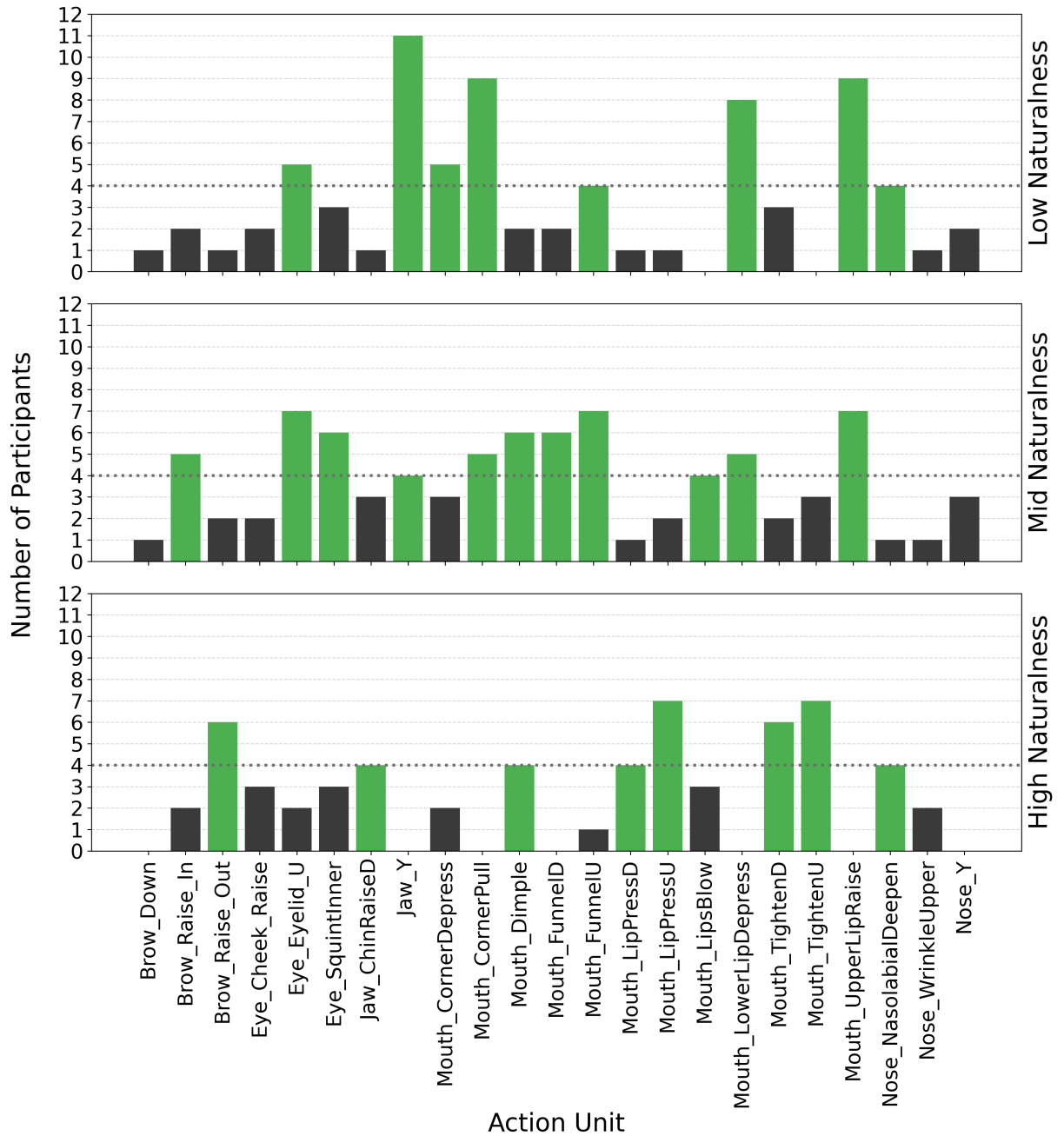


Figure 4.7: Frequency of facial action units (AUs) significantly associated with perceived naturalness across participants. Bar plots show the number of participants (out of 18) for whom each AU was significantly positively correlated with naturalness ratings ($p < .05$), grouped by rating category: low (1-2), mid (3-5), and high (6-7). AU labels correspond to the naming conventions of the MetaHuman facial rig controls used to drive animation. AUs were treated as bilateral, such that if either the left or right lateralised variant of an AU was significant for a given participant, the AU was counted as significant overall. Dashed lines indicate the threshold of 4 participants (22%), which was used to determine AU inclusion in subsequent animation synthesis models.

4.3.3 Head and Eye Motion

While the initial reverse correlation analysis (described above) revealed consistent relationships between AU activations and perceived naturalness, the same was not observed for head and eye motion parameters. Regression models for these features showed mostly flat relationships with naturalness, suggesting little to no influence on participant ratings within the original dataset. However, informal post-experiment conversations revealed that many participants reported noticing head and eye movements and felt they contributed to their overall ratings. This apparent disconnect suggests that head and eye motion may play a subtler role, potentially overshadowed by the more salient facial expression changes driven by AUs.

To isolate and better model the perceptual contribution of head and eye movement, we conducted an additional condition with a small group of new participants ($n = 4$), using the same stimuli structure and procedure as the pre-validation portion of the experiment but excluding facial AUs and conducting 750 trials instead of 1500. This allowed us to examine naturalness judgments in response to head and eye motion alone, without the potentially dominating influence of facial expression cues. This additional data showed more prominent relationships between the different curve generation parameters across head, eye, and blink motion in relation to naturalness ratings and the obtained motion generation parameters across naturalness levels were used to synthesise the saccade, blink, and head motion elements of validation stimuli.

The head motion, eye motion, and blink component was collected using fewer participants and trials than the facial AU reverse correlation because it involved a substantially smaller parameter space. Unlike the facial AU dataset, which simultaneously varied a large number of facial movements, only a small number of head, eye, and blink parameters were manipulated, reducing the complexity of the stimulus space while still allowing adequate sampling of the parameter combinations.

4.3.4 Validation

To assess whether the facial animations synthesised from the reverse correlation data accurately reflected intended naturalness levels, a separate group of naïve observers ($n = 10$) rated the final animations generated from the regression models of AUs, and head and eye motion. Stimuli were created using AU subsets and temporal parameters derived from the earlier analysis and presented at multiple naturalness levels along with their corresponding head motion, saccades, and blinks. More specifically, AUs with low naturalness (scores 1-2) were combined with head and eye motion models from score 1, AUs with mid naturalness (scores 3-5) were combined with head and eye motion models from score 4, and AUs with high naturalness (scores 6-7) were combined with head and eye motion models from score 7.

Validation performance was quantified using congruence accuracy, defined as the proportion of trials in which an observer's rating category matched the intended naturalness category of the animation (Low, Mid, High). Because three response categories were available, chance performance was 33.3%. One-sample t-tests were conducted for each category to determine whether congruence accuracy was significantly greater than chance across the group.

Low naturalness animations were correctly categorised with a mean accuracy of 49.68% ($SD = 18.0\%$), which was significantly above chance: $t(9) = 2.72, p < .05$. Mid naturalness animations were categorised with a mean accuracy of 38.99% ($SD = 17.51\%$). This did not differ significantly from chance, $t(9) = 0.97, p = .179$. High naturalness animations were categorised with a mean accuracy of 62.66% ($SD = 20.49\%$), which was significantly above chance, $t(9) = 4.29, p = .001$. See *Figure 4.8* for a visualisation of the validation results. Also, see *Figure 4.9* for a confusion matrix showing the percentage of classifications for each intended naturalness category. Although the mid naturalness category showed the greatest ambiguity, these models were more frequently classified as low naturalness (42.3%) than high naturalness (18.7%), suggesting that medium naturalness was perceived as closer to the low than the high category.

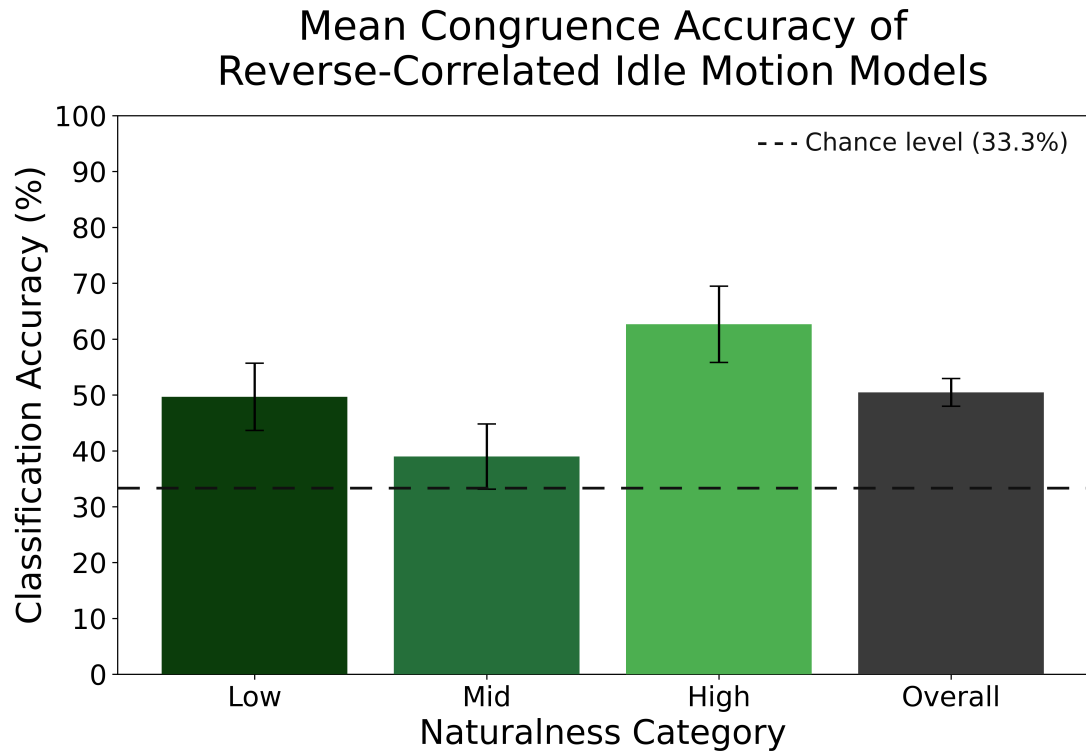


Figure 4.8: Mean congruence accuracy of reverse-correlated idle motion models across naturalness categories. Bars show the mean classification accuracy (%) with which naïve observers ($n = 10$) correctly identified the intended naturalness category (Low, Mid, High) of the generated animations, along with the overall accuracy across all conditions. Congruence accuracy was defined as the proportion of trials in which an observer's rating category matched the intended category. Error bars represent the standard error of the mean (SEM) across participants. The dashed line indicates chance performance (33.3%) corresponding to three response categories.

Confusion Matrix of Validation Judgements

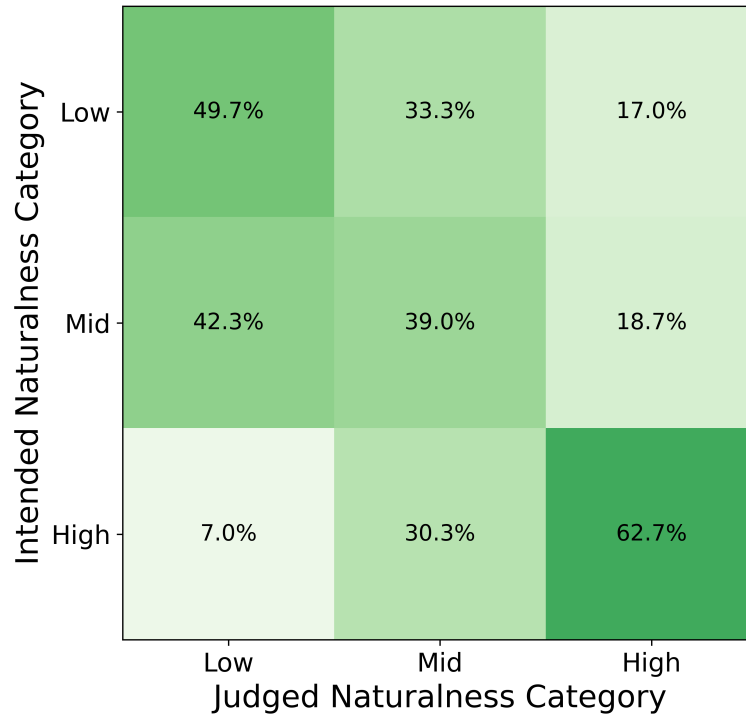


Figure 4.9: Confusion matrix showing the percentage of participant judgements for each naturalness category. Diagonal values indicate correct classifications, whereas off-diagonal values indicate misclassification across categories.

During exploratory analysis of the validation data, it became apparent that, although congruence accuracy for the medium naturalness category did not exceed chance level overall, several individual mid naturalness animations were still classified with high accuracy by naïve observers. Examining the most consistently rated animations therefore provides additional insight into within-category variability in perceptual clarity (see *Table 4.1* & *Figure 4.10*). Notably, overall the most accurately classified animations spanned multiple characters across all three naturalness levels, and some characters appeared among the most accurate examples at more than one level. This pattern suggests that observers' judgements were not driven by a particular character's appearance but were instead sensitive to differences in motion characteristics across animations, in-line with the requirements of the experimental task.

Intended Naturalness	Video ID	Congruence
High	42_kioko.mp4	90%
	84_wallace.mp4	90%
	23_irene.mp4	80%
	53_layla.mp4	80%
	60_myles.mp4	80%
Mid	22_irene.mp4	70%
	34_kendra.mp4	70%
	52_layla.mp4	70%
	03_bernice.mp4	60%
	09_cooper.mp4	60%
Low	07_cooper.mp4	90%
	02_bernice.mp4	80%
	19_irene.mp4	80%
	49_layla.mp4	80%
	50_layla.mp4	80%

Table 4.1: *The five validation animations within each intended naturalness category (Low, Mid, High) that were most consistently classified correctly by naïve observers. Congruence represents the percentage of observers whose ratings matched the intended category of the animation.*



Figure 4.10: Representative frames from the three most consistently classified validation animations in each intended naturalness condition (High, Mid, Low), shown at approximately peak facial action unit activation. Animations are listed in order of appearance. Percentages indicate the proportion of naïve observers whose ratings matched the intended naturalness category of each animation. High naturalness: 84_Wallace (90%), 42_Kioko (90%), 23_Irene (80%). Mid naturalness: 52_Layla (70%), 34_Kendra (70%), 22_Irene (70%). Low naturalness: 02_Bernice (80%), 07_Cooper (90%), 19_Irene (80%).

4.4 Discussion

The present study aimed to use a reverse correlation approach to identify the patterns of subtle, non-expressive idle facial motion that drives perceptions of naturalness. Specifically, the study focused on whether perceived naturalness in neutral facial behaviour could be modelled in a data-driven manner, and whether such models could be used to synthesise novel animations that

naïve observers would perceive at intended levels of naturalness.

To address this, a large-scale reverse correlation experiment was conducted using procedurally generated idle facial animations applied to Unreal Engine’s MetaHuman characters. Facial action units (AUs), along with head motion, eye saccades, and blinks, were systematically varied across a large stimulus set, and participants rated the naturalness of each animation. Reverse correlation analyses were then used to identify relationships between motion generation parameters and perceived naturalness, and to derive generative models of idle facial behaviour across the different naturalness levels. These models were subsequently validated in a separate phase of the experiment in which naïve observers rated new animations synthesised directly from the resulting reverse correlation models of natural idle motion. The reverse correlation analysis revealed systematic relationships between facial AU activations and head/eye motion parameters, and perceived naturalness. Distinct sets of AUs were associated with low, mid, and high naturalness ratings, with some AUs showing consistent associations across a large proportion of participants. Formal statistical tests confirmed that the prevalence of several key AUs differed significantly across naturalness levels, including jaw opening and upper lip raising actions that were significantly more common in low naturalness animations than in high naturalness animations. These relationships between AUs and head/eye motion, and naturalness ratings were used to generate new animations at three intended levels of naturalness: low, medium, and high.

Validation results revealed that naïve observers categorised low and high naturalness animations with accuracy significantly above chance levels, demonstrating that the reverse correlation models successfully captured motion patterns that reliably expressed differences in perceived naturalness of idle facial motion. Medium naturalness animations, while not categorised above chance levels overall, showed substantial variability, with several individual animations classified accurately by the majority of observers. Together, these findings demonstrate that reverse correlation can be applied to idle facial motion to recover perceptually meaningful movement patterns, and that these patterns can be used to synthesise new animations with predictable effects on perceived naturalness.

To better understand what distinguished the validated low and high naturalness animations, it is informative to examine the patterns of facial AU activation revealed by the reverse correlation analysis. Additional statistical testing confirmed that several key lower-face AUs differed significantly in prevalence across naturalness levels, particularly Jaw Drop (AU26), Upper Lip Raise (AU10), and Lip Corner Pull (AU12), which were significantly more common in low naturalness animations than in high naturalness animations. In terms of the reverse correlation analysis, low naturalness was predominantly associated with AUs that produce relatively large, salient facial movements, particularly in the lower face. These included mouth opening and jaw related actions such as Upper Lip Raise (AU10), Jaw Drop (AU26), and lip corner movements (e.g., AU12, AU16), which often expose the teeth or noticeably alter the mouth shape. When

expressed in an idle context, such movements are likely to violate observers' expectations for subtle, neutral facial behaviour, contributing to impressions of unnaturalness. Whereas, high naturalness was characterised by a different pattern of AU activity, involving more constrained and localised movements, such as lip pressing and tightening (e.g., AU24, AU23) and subtle brow-related actions (e.g., AU2). These AUs tend to produce smaller changes in facial configuration and are more consistent with the low-amplitude, non-communicative movements typically expected in natural idle facial behaviour. Together, these patterns suggest that perceived naturalness in idle faces is not driven by the presence of movement per se, but by the scale and functional plausibility of that movement within a neutral facial expression context.

Due to the exploratory nature of the present study into idle facial motion, more advanced analyses examining the relationship between AU combinations and perceived naturalness were not carried out. Specifically, the analysis considered each AU independently and did not examine the co-occurrence of multiple AUs or how interactions between simultaneously active facial movements may influence naturalness judgements. Thus, the reported findings should be interpreted as describing the individual associations between AU occurrence and perceived naturalness rather than the combined effects of specific facial movement patterns. An important next step would be to jointly model AU occurrence using approaches such as logistic regression or mixed-effects models, which would allow the effects of co-occurring AUs on perceived naturalness to be identified.

Furthermore, although the combined naturalness models were successfully validated, the head motion, eye motion, and blink parameters incorporated into these models were derived from a separate reverse correlation dataset involving fewer participants than the facial AU component. Thus, the final animations implicitly treated the independently derived head/eye and facial components with equal weighting. Future work should attempt to also jointly model facial AUs, head motion, eye motion, and blink behaviour, allowing the contribution of each motion component towards perceived naturalness to be established rather than treating them as equally important.

4.4.1 MetaHumans and Reverse Correlation

Alongside the empirical findings, the present study makes a methodological contribution by demonstrating that reverse correlation can be successfully implemented using state-of-the-art high-fidelity digital human avatars. Integrating MetaHuman facial rigs (Epic Games, 2025) with a custom procedural animation pipeline in Unreal Engine (see *Chapter 2*) enabled large-scale, perception-driven modelling of facial motion using state-of-the-art production-grade characters rather than the lower graphical fidelity avatars used in psychological research to date (Jack & Schyns, 2017; Todorov et al., 2025). This framework supports an end-to-end reverse correlation

workflow: from automated stimulus generation and rendering to perceptual evaluation and data-driven synthesis, while allowing high volume and variability in facial motion without reliance on manual animation or large motion-capture datasets. Thus, the study establishes the feasibility of applying reverse correlation methods to contemporary animation and game-engine pipelines without compromising the necessary experimental control required by reverse correlation approaches.

This methodological contribution also builds directly on the previous chapter (Chapter 3) validating MetaHumans as suitable stimuli for perceptual research and extends prior reverse correlation work that has largely relied on static images or dynamic faces with lower graphical fidelity digital human characters (Jack & Schyns, 2017; Todorov et al., 2025). While the present framework is not intended to inevitably replace existing reverse correlation approaches, it demonstrates that such methods can be successfully applied to modern, high graphical fidelity digital human characters. Thus, the current study serves as a “proof of concept” for conducting perceptually grounded, data-driven modelling of human behaviour using state-of-the-art solutions employed across various industries.

Another contribution of the present study is the extension of applying reverse correlation beyond its traditional focus on overtly expressive facial behaviour. To date, most reverse correlation work in face perception has targeted clearly interpretable social signals, including emotional expressions, personality traits, and communicative mental states such as trustworthiness, dominance, confusion, or interest (Chen et al., 2015; Dotsch & Todorov, 2012; Gill et al., 2014; Jack & Schyns, 2017; Todorov et al., 2025). In such studies, the examined facial movements are typically salient, intentional, and strongly linked to specific meanings that observers can categorise. Extending this line of work, the present study demonstrates that reverse correlation can also be applied to more subtle and neutral, non-communicative facial behaviour.

To evidence this, rather than asking observers to judge emotions or traits, participants focused on perceiving subtle idle facial motion that occurred in the absence of explicit expression or social intent. Despite this lack of overt signalling, the reverse correlation analysis revealed systematic relationships between specific facial action units and perceived naturalness, indicating that idle facial behaviour is not simply perceptually meaningless noise. Instead, even subtle background facial movements carry structured information that observers use when forming impressions of whether a face appears alive and natural. Just as observers are sensitive to subtle idle movement in the body (Landa et al., 2025; Treal et al., 2020, 2021), based on our results they are also sensitive to subtle idle movement in the face itself, even in the absence of overt facial expression.

Together, these findings challenge the common assumption, inherent to some animation practice and perceptual research, that idle motion is often conceptualised as simple filler or vi-

sual padding between meaningful actions (Bates, 1994; Kocoń, 2013). While idle behaviours are often added based on animator intuition or practical convention to prevent characters from appearing static, the present results suggest that observers are sensitive to distinct patterns of idle facial motion, and that these patterns can be recovered using data-driven perceptual methods. Importantly, the validation phase showed that animations generated from these reverse correlation models could reliably convey differences in perceived naturalness of idle motion to naïve observers, even without any explicit emotional or communicative cues.

4.4.2 Perceived Naturalness of Idle Facial Motion

The validation results provide strong evidence that the reverse correlation models successfully captured perceptually salient patterns of idle facial motion at the extremes of the naturalness scale. Animations synthesised from the low and high naturalness models were classified by naïve observers with accuracy significantly above chance, indicating that these models did not simply describe statistical regularities in the reverse correlation data but were able to generate motion that reliably conveyed intended perceptual categories.

Importantly, this effect was observed across multiple characters rather than being driven by a small subset of stimuli or identities. The consistency of low and high naturalness classifications across different faces suggests that the extracted motion patterns generalised beyond individual characters and reflected shared perceptual expectations about what constitutes unnatural versus natural idle facial behaviour. This supports the claim that the reverse correlation approach recovered motion features that observers consistently associate with naturalness judgments, rather than particular artefacts or features tied to specific avatars.

At the level of facial action units, this robustness aligns with the observed AU patterns. Low naturalness was associated with relatively large, salient movements, particularly in the lower face, that are atypical of idle behaviour, whereas high naturalness involved more constrained, localised actions consistent with subtle background facial dynamics. Together, the validation congruence results and AU consistency indicate that the low and high naturalness models captured stable and perceptually meaningful structure in idle facial motion. As such, these models can be viewed as generative and perceptually valid: they are capable of producing novel idle facial animations that observers reliably interpret in line with their intended naturalness level.

In contrast to the robust performance observed for low and high naturalness, medium naturalness animations were not classified above chance level when considered as a category overall. Rather than indicating a failure of the modelling approach, this result is theoretically informative and highlights an important property of perceived naturalness of idle facial motion.

Naturalness is unlikely to be a strictly categorical dimension, particularly in the middle

of the scale. Whereas, highly unnatural motion may clearly violate perceptual expectations and highly natural motion may clearly conform to them, mid-range motion is inherently more ambiguous. Animations in this range may contain competing cues, some more consistent with natural idle behaviour and others deviating from it, leading different observers (or even the same observer across trials) to perceive the motion in different directions of the naturalness spectrum.

This interpretation is supported by the validation data at the level of individual stimuli. Despite the absence of a significant effect at the category level for mid naturalness, several mid naturalness animations were still classified correctly by a majority of observers. This suggests that some combinations of facial action units and temporal parameters communicate “mid” naturalness more clearly than others. In the present study, however, AU subsets for validation animations were sampled randomly from the pool of significant AUs associated with medium naturalness. While this ensured that the validation stimuli were novel and varied, it also likely diluted category clarity by combining AUs that may differ in how strongly they signal intermediate naturalness.

Importantly, the aim of the validation phase was not to optimise classification accuracy, but to provide a ‘first pass’ test of whether reverse correlated models of idle facial motion could generalise to new animations and be interpreted meaningfully by naïve observers. From this perspective, the observed ambiguity in the mid naturalness condition is not a limitation of the approach but a reflection of the underlying perceptual continuum. It points to a promising direction for future work: systematically examining which specific AU combinations or motion profiles most reliably convey the different levels of naturalness, and whether constraining generation to these subsets can enhance perceptual clarity across all naturalness levels.

Taken together, the validation results suggest that reverse correlation is well suited to capturing the extremes of perceived naturalness, while also revealing the potential graded and context-dependent nature of mid-range idle facial motion. This reinforces the view that naturalness is best understood as a continuous perceptual dimension rather than a set of discrete categories, and that ambiguity in the middle of the scale is a meaningful outcome rather than an experimental shortcoming.

Consistent with this interpretation, examination of the validation data at the level of individual stimuli further indicates that observers’ judgements were driven by motion rather than by character appearance. The most accurately classified animations within each naturalness level were distributed across multiple digital human characters, and several characters appeared among the most accurately classified stimuli at more than one level of naturalness. This pattern suggests that perceived naturalness varied systematically with the applied motion profiles rather than being tied to specific facial identities. Together with the graded nature of the validation results, this stimulus-level evidence reinforces the view that observers evaluated idle facial motion

along a continuous perceptual dimension, attending to how movement patterns unfolded over time rather than relying on static appearance cues.

4.4.3 Implications for Uncanny Valley Research

A central implication of the present findings relates to their relevance for research on the uncanny valley. To date, much of the uncanny valley literature has focused on the role of visual appearance, particularly realism of facial features, skin texture, and overall human likeness, in shaping affective responses such as eeriness and discomfort (Di Natale et al., 2023; Kätsyri et al., 2015; Mori, 1970; Mori et al., 2012). While motion has long been acknowledged as a contributing factor, it has often been treated as secondary to appearance when studied and in general is researched much more sparsely, making it difficult to isolate its specific contribution to uncanny responses.

The present study may help address this limitation by providing a framework in which the naturalness of idle facial motion can be manipulated systematically while holding visual appearance constant. This separation of motion from appearance enables a more controlled examination of whether, and how, motion alone contributes to the uncanny valley effect, responding directly to calls in the literature for more precise experimental control over dynamic cues (Di Natale et al., 2023; Diel et al., 2023; Kätsyri et al., 2015).

Saygin et al. (2012) provide neural evidence in line with mismatch based accounts of the uncanny valley, reporting distinctive neural responses in brain regions involved in perceiving and interpreting action when an android appeared human but moved non-biologically. This interpretation aligns with Kätsyri et al., (2015) and Di Natale et al., (2023), who argue that uncanny responses are often driven by perceptual mismatch or inconsistency, including cases where an entity's behaviour does not fit its humanlike appearance. However, Saygin et al.'s (2012) manipulation was categorical (biological vs. mechanical), leaving open which specific motion cues matter. The present data-driven approach addresses this by identifying perceptually salient features of facial motion naturalness and using them to guide motion synthesis, rather than relying on broad labels or intuition.

By validating low, high, and to an extent medium naturalness motion profiles, the current framework enables a well controlled test of motion-based uncanny valley effects in which appearance realism can be held constant and motion naturalness varied in a graded, perceptually grounded manner. The present results therefore provide the methodological foundation for examining how subtle changes in idle facial motion affect the measures related to the uncanny valley effect, including eeriness, humanness, realism, and familiarity (as was the case in *Chapter 3*: comparison between GFG and MetaHuman characters).

These considerations directly motivate the final study of this thesis. Building on the validated reverse correlation models, the next experiment systematically applies validated low, medium, and high naturalness idle facial motion to MetaHuman characters with consistent levels of graphical fidelity and measures observers’ responses along established uncanny valley dimensions. This approach positions motion naturalness as a precisely controlled independent variable, allowing a more specific test of the role of motion in shaping affective and social evaluations of digital humans.

4.4.4 Applied and Practical Implications

Beyond its theoretical contributions, the present work has many practical implications for the creation and evaluation of facial motion in digital characters. Specifically, the findings relate to ongoing challenges in animation practice and interactive media production, where generating believable idle behaviour remains difficult to scale and standardise. By grounding idle motion synthesis in perceptual data rather than animator intuition or motion capture alone, the current framework suggests an alternative approach to generating subtle facial dynamics that are perceptually meaningful while remaining computationally feasible.

Animation and Motion Synthesis

In contemporary animation pipelines, idle motion is often authored manually or derived from motion capture. In many cases, decisions about idle behaviour are guided by animator experience, stylistic convention, or production constraints, rather than by systematic evidence about how observers perceive subtle dynamics of idle motion. As a result, the quality and consistency of idle behaviour can vary, particularly when such motion is reused across characters or deployed at scale.

Motion capture provides an accurate, high-fidelity alternative to handcrafted idle animation, however, producing idle motion data is often costly and publicly available 3D idle datasets remain scarce (Landa et al., 2025). Landa and colleagues argue that capturing genuinely natural idle behaviour is often seen as difficult since awareness of recording may alter movement and motion capture suits can be intrusive. However, their results suggest that this constraint may be overstated, as observers were not able to reliably distinguish acted from genuine idle motion. Notably, observers could distinguish recorded idle motion (genuine or acted) from handmade idle animations, indicating that handcrafted idles are perceptually separable from motion-captured idle movement patterns.

The present framework offers a complementary approach to both hand-crafted animation

and motion capture by using reverse correlation to derive generative models of idle facial motion directly from perceptual judgments. Rather than aiming to reproduce realistic motion trajectories manually, the approach targets perceptually salient patterns that reliably influence observers' impressions of naturalness. In this sense, the generated motion is grounded in how movement is interpreted rather than how it is practically produced. We do not claim that perception-driven synthesis can replace motion capture or skilled animation, particularly for expressive or narrative facial performance; instead, our results demonstrate a promising middle ground for modelling neutral and background facial behaviour, where the goal is not expressive communication but the maintenance of a convincing sense of aliveness. Thus, the framework may be particularly relevant for applications where large volumes of subtle facial motion are needed and where consistency and controllability are prioritised over bespoke performance.

The relevance of this framework becomes especially apparent in large-scale interactive environments such as contemporary video games, virtual worlds, and simulations, where the need to animate many characters places practical constraints on bespoke facial animation. In these contexts, a common challenge is the disparity in motion quality between main characters, who receive substantial animation resources, and side or background characters, whose idle facial behaviour may be drastically less polished and realistic. Procedural animation systems are often used to address this scalability problem, but they typically prioritise efficiency over perceptual subtlety, resulting in idle motion that can appear overly uniform, exaggerated, or misaligned with observers' expectations of natural behaviour. The present findings suggest that even small deviations in idle facial dynamics can influence perceived naturalness, highlighting the importance of modelling such behaviour in a perceptually grounded way. By providing a data-driven and parametric method for synthesising idle facial motion, the current framework points toward a way of improving behavioural plausibility across large character populations without relying on extensive manual animation or motion capture. While exploratory rather than production-ready, this work demonstrates the potential for perception-driven modelling to inform character animation pipelines in domains where scalable and perceptually plausible idle behaviour is needed.

4.4.5 Future Directions

The present framework raises several open questions for future work. One line of inquiry concerns how specific combinations of facial AUs work in conjunction with temporal parameters to shape perceived naturalness across the scale. Although mid-range idle motion was classified accurately with less consistency overall, some individual stimuli were judged reliably with high consistency, suggesting that certain AU configurations convey medium naturalness more clearly than others. Similarly, by investigating how AU combinations impact perception of the low and high naturalness models, the classification accuracy could potentially be boosted there as

well. Future work could therefore move beyond random sampling of AU combinations to better characterise how different regions of the naturalness continuum are defined perceptually.

A second avenue involves extending reverse correlation beyond facial dynamics to full-body behaviour. Having demonstrated that high-fidelity MetaHuman characters can be successfully used in reverse correlation experiments, future work is no longer limited to head-only rigs, as was the case with earlier tools such as GFG or FaceGen (Jack & Schyns, 2017; Todorov et al., 2025). Access to full-body rigs enables the systematic manipulation of torso, head, and limb movements, which are known to affect how we perceive digital humans (Landa et al., 2025; Treval et al., 2020, 2021). Applying perceptual modelling techniques at the body level would therefore allow investigation of how facial and bodily cues jointly shape judgments of naturalness and perception more broadly.

Future work could also leverage the parametric nature of MetaHuman character models to investigate how systematic variation in body morphology shapes human perception. Prior research has shown that facial structure and morphology are tightly linked to trait inferences such as attractiveness, dominance, trustworthiness, and more (Gill et al., 2014; Gosetti et al., 2023, 2024; Oosterhof & Todorov, 2009; Zhan et al., 2020, 2021), yet similar questions remain underexplored at the level of the body. Reverse correlation has previously been used to visualise mental representations of body shape using noise distorted two-dimensional body silhouettes (Lick et al., 2013; Maister et al., 2021). MetaHuman characters provide fine grained control over a wide range of body shape parameters, including musculature, fatness, overall body proportions, and more, allowing large sets of systematically varied bodies to be generated at scale. This opens the possibility of applying reverse correlation methods to body appearance itself, using perceptual judgments to identify which morphological features drive trait attribution or aesthetic preference in detailed three-dimensional body models.

More broadly, the accessibility of MetaHuman characters also enables important opportunities for replication and methodological comparison in social perception research. Because MetaHumans are freely available, they lower barriers to reproducing classic and contemporary findings using high-fidelity digital humans, which is particularly valuable in light of ongoing concerns about replicability in psychological science (Shrout & Rodgers, 2018). Prior work suggests that character appearance can meaningfully shape how facial expressions and social signals are perceived (Higgins et al., 2021), raising the question of whether results obtained using earlier platforms with lower graphical fidelity generalise to more realistic digital humans. Systematically revisiting established findings with MetaHuman characters would allow researchers to test the robustness of these effects and assess how advances in visual fidelity interact with perceptual and social judgments. In this sense, the present framework not only supports new experimental questions but also provides a practical foundation for replication and validation of existing theories using contemporary, high-resolution character models.

4.5 Conclusion

This chapter demonstrates that idle facial motion carries structured information that observers reliably use when judging the naturalness of a neutral digital face. Using a reverse correlation approach, the present work shows that subtle patterns of facial AU activation and temporal dynamics systematically relate to perceived naturalness, even in the absence of overt expression or communicative intent. Consequently, the study also establishes the feasibility of applying reverse correlation paradigms to contemporary, production-ready MetaHuman characters (Epic Games, 2025). Taken together, these findings extend prior reverse correlation research beyond expressive and trait-based judgments, establishing idle facial behaviour itself as a salient perceptual signal.

Importantly, the resulting models do more than simply describe associations between facial motion and perceived naturalness. Animations synthesised from the low and high naturalness models were consistently interpreted by naïve observers in line with their intended naturalness levels. This demonstrates that the reverse correlation models can be used to generate new idle facial animations that produce predictable perceptual outcomes, rather than only summarising patterns in the original data. At the same time, the less consistent classification of medium naturalness animations indicates that perceived naturalness is not a sharply defined category. Instead, naturalness appears to be a graded perceptual dimension, with greater ambiguity in the middle of the scale. Thus, the present work demonstrates that reverse correlation can be used to derive generative models of idle facial motion with reliable and testable perceptual consequences.

By validating a set of perceptually grounded idle facial motion profiles while holding visual appearance constant, the present chapter establishes a methodological foundation for isolating motion as an independent factor in social perception. The final empirical chapter builds directly on this framework by applying the validated low, medium, and high naturalness animations to MetaHuman characters and examining how systematic variation in idle facial motion naturalness shapes uncanny valley responses, including eeriness, humanness, realism, and familiarity.

Chapter 5

Motion Naturalness and the Uncanny Valley in Digital Avatars (Study 3)

5.0.1 Chapter Overview

This chapter advances the central aim of the thesis: understanding how digital humans come to feel alive rather than uncanny by isolating behavioural realism as a key driver of social evaluation. Building on the validated MetaHuman platform and reverse-correlated models of idle motion developed in earlier chapters, the present study examines the mismatch-based accounts of the uncanny valley by holding appearance constant and manipulating perceptually derived motion naturalness. The findings show that unnatural idle motion reliably increases eeriness and reduces humanness, familiarity, and realism relative to natural motion, whereas natural motion largely preserves baseline impressions rather than dramatically enhancing them. This asymmetric pattern supports the broader thesis argument that uncanny responses are driven more by violations of perceptual coherence than by realism itself. Methodologically, the chapter demonstrates a full pipeline: from high-fidelity avatars to perceptually grounded motion models, to causal evaluation of affective outcomes, while demonstrating that avoiding mismatches between appearance and dynamics may be more critical for lifelike digital characters than simply adding greater behavioural complexity.

5.1 Introduction

5.1.1 The Uncanny Valley: From Appearance to Perceptual Mismatch

The uncanny valley was first proposed by Mori (Mori, 1970; Mori et al., 2012), who suggested that as artificial agents become increasingly humanlike, observers' affinity increases until a critical point at which near-human entities induce discomfort or eeriness. In this original formulation, the valley was primarily framed in terms of appearance realism, with visual proximity to the human form seen as the key driver of negative affect. Early empirical work therefore focused heavily on manipulating degrees of facial realism to test this prediction (e.g., MacDorman et al., 2009; Seyama & Nagayama, 2007).

However, subsequent research has shown that realism alone does not inevitably lead to elevated eeriness. Highly realistic digital humans can be perceived as more humanlike and less eerie than lower fidelity counterparts (Higgins et al., 2021), and broader reviews emphasise mixed support for a simple realism-eeriness curve (Di Natale et al., 2023; Kätsyri et al., 2015). These findings have shifted theoretical focus towards perceptual mismatch accounts, which propose that discomfort arises when cues that are typically coherent in real humans, such as physical appearance and facial expression, become inconsistent.

This interpretation aligns with predictive processing frameworks, which argue that observers generate expectations about biological motion and experience discomfort when those expectations are violated (Diel & Lewis, 2022). However, similar effects can also be explained by bottom-up perceptual mechanisms, whereby observers attempt to classify incoming visual information and experience unease when the stimulus deviates from familiar biological patterns. Experimentally introducing subtle temporal asynchronies into facial animation reliably increases uncanniness across human and artificial agents (Diel et al., 2023). Converging neural evidence further supports this view: androids with humanlike appearance but non-biological motion elicit distinctive activation in action-perception regions of the brain, which may reflect either violations of predicted movement dynamics or the perceptual detection of atypical motion cues (Saygin et al., 2012).

Together, these findings suggest that uncanny responses are driven less by realism of itself than by inconsistencies between expected and observed human cues. Yet despite the central role of motion in these accounts, dynamic cues remain relatively understudied experimentally.

5.1.2 Motion as a Critical but Under-Isolated Component of the Uncanny Valley

If uncanny responses arise from perceptual mismatch, motion should be one of the central determinants of affective evaluation. Mori (1970) himself speculated that movement would amplify the valley effect, and contemporary mismatch accounts suggest that dynamic coherence is likely one of the core aspects of the phenomenon (Di Natale et al., 2023; Kätsyri et al., 2015). Importantly, based on the reasoning behind mismatch accounts, motion may either increase affinity or intensify uncanniness, depending on its biological plausibility. Subtle temporal asynchronies in facial animation reliably increase perceived eeriness across human, android, and virtual faces (Diel et al., 2023), supporting the view that violations of expected motion dynamics drive discomfort (Diel & Lewis, 2022).

Likewise, biologically plausible motion appears to enhance perceived acceptability and engagement. Characters animated with natural movement are typically judged as more acceptable and humanlike than those with artificial dynamics (Piwek et al., 2014), and natural human body motion captures visual attention even independently of appearance (Williams et al., 2019). At a more granular level, adding subtle postural oscillations increases the believability and perceived intensity of facial expressions in avatars (Treal et al., 2020, 2021), while introducing idle motion during otherwise inactive moments increases perceived animacy and friendliness in humanoid robots (Asselborn et al., 2017). Together, these findings suggest that ongoing, low-level movement functions as a cue to aliveness and appears to have a large influence on how avatars may be perceived.

Despite this theoretical relevance, motion remains understudied empirically. Many studies manipulate broad categories (e.g., biological vs mechanical motion; Saygin et al., 2012), rely on lower-fidelity characters, or vary motion and appearance simultaneously, making it difficult to determine whether observed effects are driven by dynamics, rendering, or their interaction. Furthermore, idle behaviour defined as subtle, non-expressive background motion, has only recently been recognised as both perceptually impactful and understudied in itself (Landa et al., 2025).

To isolate the role of motion within the uncanny valley framework, two methodological requirements are necessary: firstly, a validated high-fidelity appearance platform that avoids confounding realism with motion; and second, a perceptually grounded manipulation of motion quality itself.

5.1.3 Establishing a High-Fidelity Baseline: MetaHumans as Controlled Stimuli (Chapter 3)

Before motion effects can be precisely isolated, it is necessary to ensure that the underlying characters are not inherently uncanny. Reverse correlation research in social face perception has demonstrated the value of highly controlled, parametric avatar systems for uncovering perceptually meaningful facial dynamics (Jack & Schyns, 2017; Todorov et al., 2011). Platforms used in this line of research, such as FaceGen (Oosterhof & Todorov, 2009) and GFG (Yu et al., 2012), enable systematic manipulation of facial structure and movement, making them ideal tools for modelling internal perceptual representations. However, their graphical fidelity reflects the technological standards of the time in which they were developed, making them inferior to current state-of-the-art solutions in terms of graphical fidelity. Importantly, this limitation mainly concerns appearance: the GFG tool models facial motion using parameterised dynamic functions that describe how action unit movements unfold over time.

This distinction is important. If an avatar's base appearance is limited by rendering quality, topology, or shading, perceptual responses may reflect, or be distorted by, artefacts of graphical fidelity rather than specific reactions to motion or cue coherence (Crookes et al., 2015; McDonnell et al., 2012). In such cases, attempts to isolate motion effects on the uncanny valley phenomenon risk being confounded by pre-existing appearance-driven perceptions and impressions. Thus, although reverse correlation methodologies provide a powerful framework for modelling perceptually valid facial dynamics, applying them to uncanny valley questions requires a platform capable of supporting a certain level of character fidelity. *Study 1a (Chapter 3)* addressed this issue directly by benchmarking Epic Games' MetaHuman framework (Epic Games, 2021) against the more established GFG avatars (Yu et al., 2012) using validated uncanny valley measures (Diel et al., 2021; Ho & MacDorman, 2010, 2017).

Across measures of humanness, realism, and familiarity, MetaHumans were rated significantly higher than the GFG characters, importantly, while showing no significant increases in eeriness. This pattern indicates that increased graphical fidelity does not inevitably produce uncanny discomfort, aligning with mismatch-based accounts (Kättsyri et al., 2015) and empirical findings that high fidelity rendering can avoid the valley when perceptual cues remain coherent (Bae et al., 2025; Higgins et al., 2021).

Together, these findings establish MetaHumans as a valid, high fidelity stimulus platform that does not appear to inherently elicit strong uncanny responses. Thus, providing a stable perceptual baseline from which motion can be manipulated independently of appearance. Having validated the visual foundation, the next methodological challenge was to define and manipulate motion naturalness in a perceptually grounded way rather than relying on animator intuition,

broad categorical contrasts, or crude motion distortion.

5.1.4 Modelling Idle Motion Naturalness Through Reverse Correlation (Chapter 4)

To isolate motion as an experimental variable, it was necessary to ensure that differences in motion quality were not defined arbitrarily. Rather than contrasting broad categories such as ‘biological’ versus ‘mechanical’ movement, *Study 3 (Chapter 4)* adopted a reverse correlation approach to derive motion patterns directly from observers’ perceptual judgements. Reverse correlation has been widely used in social perception research to pinpoint the stimulus features that drive subjective impressions without imposing strong a priori assumptions (Jack & Schyns, 2017; Todorov et al., 2011). Dynamic extensions of this method have successfully modelled facial expressions by randomly varying facial action units (AUs) and linking them to emotion judgements (Jack, Caldara, & Schyns, 2012; Jack, Garrod, et al., 2012; Jack & Schyns, 2017; Jack et al., 2014; Yu et al., 2012). However, to date, no reverse correlation work has subtle, non-expressive idle facial motion.

Applying reverse correlation to procedurally generated MetaHuman animations revealed systematic patterns in perceived naturalness of idle facial motion. Distinct action unit (Ekman & Friesen, 1978) patterns were associated with low versus high naturalness ratings. Low naturalness was characterised by relatively large, salient lower-face movements (e.g., jaw drop, lip raising), which were judged as atypical of neutral idle behaviour. Whereas high naturalness involved more constrained, localised movements (e.g., subtle lip pressing and brow activity) consistent with low-amplitude background facial dynamics. These findings suggest that perceived naturalness is not driven by the presence of movement in itself, but by the scale and likelihood of that movement within a neutral context.

Importantly, the validation phase demonstrated that animations synthesised from the low and high naturalness models were classified by naïve observers at levels significantly above chance. This congruence between model intended and participant perceived naturalness of motion generalised across multiple characters, and some identities appeared among the most accurately classified stimuli at more than one naturalness level. This pattern indicates that classification was driven by motion characteristics rather than by specific facial identities.

Theoretically, these findings demonstrate that idle facial motion carries structured perceptual information relevant to impressions of animacy and natural behaviour. Methodologically, they provide a validated and perceptually grounded manipulation of motion naturalness. Together with the previously validated appearance baseline, this enables motion quality to be manipulated independently of character appearance.

5.1.5 Testing Motion Naturalness Within the Uncanny Valley Framework

Although *Study 2 (Chapter 4)* established that idle motion naturalness is perceptually structured and can be modelled from observer judgements, it did not test whether these motion differences translate into shifts in core uncanny valley dimensions. In the present study, these reverse-correlated models derived from observer judgements are used generatively to synthesise new animations representing different levels of motion naturalness. Specifically, while low and high naturalness animations were reliably distinguished, it remains unknown whether degraded (i.e., unnatural) idle motion reduces perceived humanness and realism, increases eeriness, or alters familiarity. In other words, does violating expectations of biologically plausible idle behaviour shift a character's position along the uncanny valley measures?

The present experiment addresses this gap directly. By holding appearance constant using previously validated high-fidelity MetaHumans (*Chapter 3*) and systematically varying motion naturalness using perceptually validated models (*Chapter 4*), the design provides a refined examination of the mismatch theory (Di Natale et al., 2023; Kätsyri et al., 2015). If uncanniness arises from inconsistencies between cues that normally cohere in real humans, then unnatural idle motion should disrupt perceptual coherence, potentially lowering humanness and realism ratings while increasing eeriness. On the other hand, high naturalness motion should maintain cue consistency and support impressions of aliveness, perhaps potentially elevating cue consistency and reducing uncanny impressions whilst increasing humanness and realism.

Importantly, the inclusion of a still condition to the present study introduces a further theoretical comparison. Complete absence of idle motion may reduce basic animacy cues and create a different form of mismatch: a visually human face that lacks expected low-level regulatory movement. Robotics research shows that subtle idle behaviour enhances perceived humanlikeness and friendliness (Asselborn et al., 2017), and the absence of such motion can render agents statue-like. This condition therefore allows dissociation between motion presence effects (still vs animated) and motion quality effects (low vs high naturalness), enabling a more precise examination of how motion contributes to humanness, eeriness, realism, and familiarity within the uncanny valley framework.

5.1.6 Hypotheses and Broader Theoretical Contribution

Thus, building on the validated appearance platform (*Chapter 3*) and the perceptually grounded motion models (*Chapter 4*), the present experiment tests predictions regarding how idle motion naturalness modulates uncanny valley dimensions when visual appearance is held constant.

For humanlikeness, we predict that motion quality will modulate perceived animacy. High naturalness motion is likely to elicit the highest humanness ratings, as biologically plausible idle dynamics provide continuous cues to aliveness and internal regulation (Asselborn et al., 2017; Treal et al., 2020, 2021). Low naturalness motion, on the other hand, may disrupt these cues and reduce perceived humanness. The still condition serves as a theoretical baseline. In much prior work, static faces have been the default stimulus, meaning stillness may not inherently introduce cross-channel conflict; it may simply reduce or remove dynamic cues to animacy. From this perspective, we expect that characters displaying high naturalness motion will be rated as equally or more humanlike than still characters, and that still characters will be rated as more humanlike than those displaying low naturalness motion. In other words, degraded motion is expected to reduce perceived humanness relative to stillness by introducing biologically implausible dynamics, whereas natural motion should either preserve existing impressions of coherence or potentially enhance them through the addition of consistent animacy cues. We also expect this pattern to extend to measures of realism and familiarity.

For eeriness, mismatch-based accounts provide a clearer directional prediction. We expect that characters displaying low naturalness motion will be rated as more eerie than both still characters and those displaying high naturalness motion, as exaggerated or biologically implausible idle dynamics introduce incoherent cues that violate expectations for natural human movement (Diel et al., 2023; Kätsyri et al., 2015). In contrast, stillness does not introduce conflicting dynamic information, even if it reduces perceived animacy, and therefore should, in theory, elicit less discomfort than when unnatural motion is introduced. Characters displaying high naturalness motion, by maintaining coherence between appearance and behaviour, are expected to show the lowest or comparable levels of eeriness relative to stillness.

More broadly, this experiment provides the first perceptually grounded manipulation of idle motion naturalness within a high-fidelity character framework. By combining validated MetaHuman stimuli with reverse-correlated motion profiles, it offers a clean causal test of motion-based uncanny valley effects, directly responding to calls for more precise isolation of dynamic cues in uncanny valley research (Di Natale et al., 2023; Diel et al., 2023; Kätsyri et al., 2015). In doing so, it bridges perceptual modelling of motion dynamics with affective evaluation, advancing understanding of how subtle behavioural cues shape impressions of humanness, realism, familiarity, and eeriness in digital humans.

5.2 Methods

5.2.1 Participants

All participants were recruited through the University of Glasgow’s participant pool platform and opportunity sampling via word-of-mouth and flyer advertisements. The participation criteria included being aged 18 or older, having normal or corrected-to-normal vision, being free of any: 1) lexical, reading, language, hearing atypicality (e.g., dyslexia); 2) social/emotion processing related atypicality (e.g., autism spectrum disorder, depression, anxiety); 3) any face processing atypicality (e.g., prosopagnosia). In total, 36 participants were recruited (11 male, 25 female), with the mean age of 31.17 ($SD = 9.69$). All participants met the participation criteria and none were excluded. As compensation, participants were optionally entered into an anonymised prize draw for a £50 Amazon voucher. The study received ethical approval from the University of Glasgow’s College of Medical, Veterinary & Life Sciences Ethics Committee (Ethics Code: 200240259).

5.2.2 Materials

The experiment was administered online using Qualtrics (Provo, UT), which hosted the questionnaire, presented the video stimuli, and recorded participant responses. All video clips were embedded directly within the survey interface and were required to play in full before rating scales were displayed. This ensured consistent exposure duration across trials and prevented participants from advancing without viewing the stimuli.

The stimuli consisted of six MetaHuman characters created using Unreal Engine (Epic Games, 2025). To maintain methodological continuity with the benchmarking studies presented in *Chapter 3*, character selection followed the same balanced structure. The final stimulus set comprised three male and three female characters, representing three ethnic categories: White, East Asian, and African heritage. Each ethnic category included one male and one female character (see *Fig.5.1*). This balanced design was implemented to minimise the possibility that any observed effects could be attributed to gender or ethnicity rather than to differences in motion characteristics. All characters were rendered at the highest level of visual fidelity, and lighting, camera position, background, and rendering settings were held constant across all stimuli to ensure that visual appearance remained identical between motion conditions.

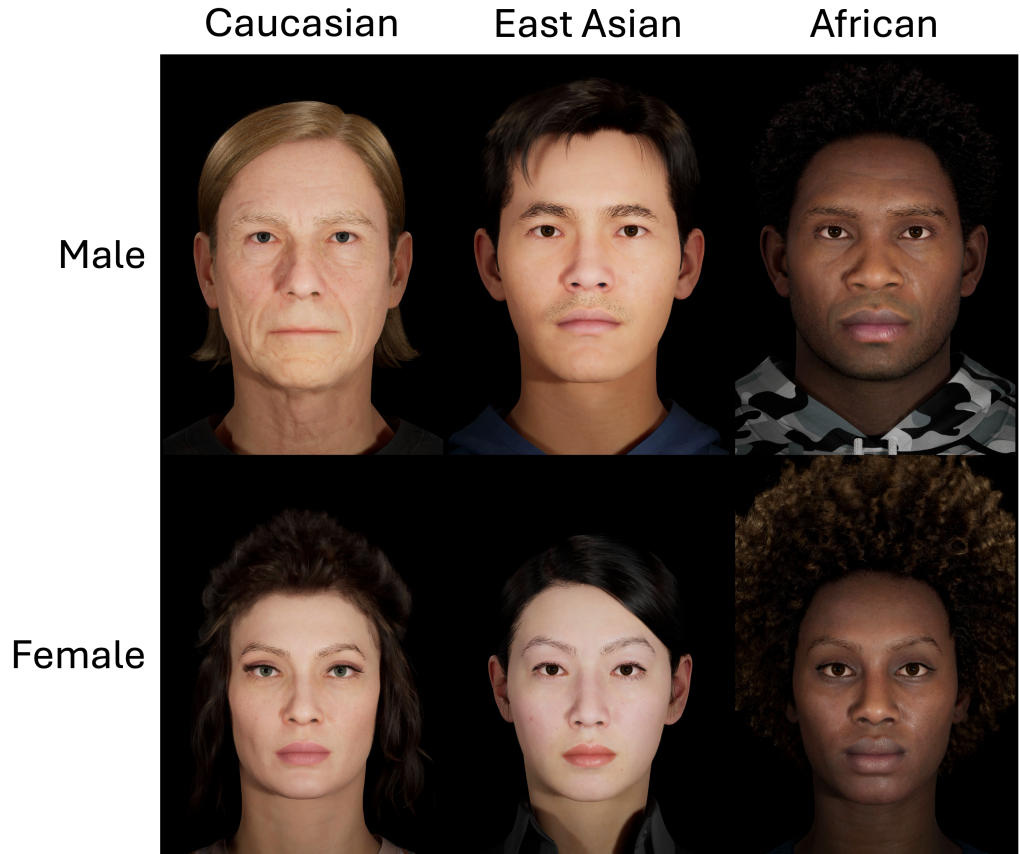


Figure 5.1: *MetaHuman characters used as base identities for the motion stimuli. The stimulus set comprised six MetaHuman characters arranged by gender (rows) and ethnicity (columns), including Caucasian, East Asian, and African identities, with one male and one female character representing each category.*

The study employed three motion conditions: a still condition, a low naturalness idle animation condition, and a high naturalness idle animation condition. All animated clips were six seconds in duration and presented in a frontal head-and-shoulders view without speech or overt emotional expressions. The animated clips contained continuous idle head motion, eye saccades, and blink activity, with a brief segment of facial action unit (AU) activity embedded midway through the clip. For each animation, a small subset of AUs was randomly sampled from the pool identified as significant for the corresponding naturalness level in the reverse correlation analysis, ensuring that each clip contained a unique but model-constrained configuration of idle facial behaviour. Full technical details of the animation generation procedure and parameter estimation are provided in *Study 2 (Chapter 4)*.

Selection of the final animated stimuli was guided by the congruence results obtained in the validation phase of the previous study. Specifically, the six most consistently classified animations were identified separately for the low and high naturalness models. These animations demonstrated the highest agreement between intended naturalness category and naïve observer ratings, thereby providing the strongest perceptual manipulation of motion quality. As reported

in the previous chapter, the most accurately classified animations spanned multiple characters and naturalness levels, and some characters appeared among the most congruent examples at more than one level. This pattern indicated that observers’ judgements were driven primarily by motion characteristics rather than by specific character identity, supporting the use of these animations as robust exemplars of low and high naturalness.

Character selection for the present study was informed by these highly congruent stimuli while ensuring balanced representation across gender and ethnicity. Where a character identity appeared among the most congruent animations at both low and high naturalness levels, that character was retained directly for the corresponding condition. In cases where the highest congruence animations did not collectively provide full coverage across the required gender and ethnic categories, the validated animation parameters from the most congruent clips were applied to additional characters using the shared MetaHuman facial rig system.

This approach preserved the validated motion characteristics while allowing identity to vary to achieve demographic balance. The specific parameter sets used to generate the animations for each character in the low and high naturalness conditions are summarised in *Table 5.1*.

Character	Gender	Ethnicity	High Naturalness Model	Congruence	Low Naturalness Model	Congruence
Wallace	Male	Caucasian	84_Wallace	90%	07_Cooper	90%
Irene	Female	Caucasian	23_Irene	80%	19_Irene	80%
Myles	Male	East Asian	60_Myles	80%	62_Robin	80%
Bernice	Female	East Asian	53_Layla	80%	02_Bernice	80%
Kioko	Male	African	42_Kioko	90%	50_Layla	80%
Kwame	Female	African	78_Rowan	80%	49_Layla	80%

Table 5.1: *MetaHuman characters and reverse-correlated animation parameter sets used to generate the high and low naturalness motion stimuli. Each character was animated using parameter sets derived from the reverse correlation models reported in Study 2 (Chapter 4). The identifiers listed in the High Naturalness Model and Low Naturalness Model columns correspond to the original animations that achieved the highest classification congruence between the intended naturalness category and naïve observer ratings in the validation phase of the previous study. Congruence values indicate the percentage of observers who correctly classified the animation as belonging to the intended naturalness category. Where necessary to achieve balanced representation across gender and ethnicity, the animations derived from these validated parameter sets were applied to different MetaHuman identities, preserving the motion characteristics while allowing character identity to vary.*

The still condition served as a baseline control in which no facial animation was applied. Head pose, gaze direction, and facial expression remained neutral throughout the six-second clip. Inclusion of this condition enabled examination of whether the presence of motion itself,

independent of its perceived quality, influenced ratings of eeriness, human-likeness, familiarity, and realism.

To summarise, the animations used in the present study were selected from the most congruent examples identified during the validation phase of Study 2 (*Chapter 4*). These animations showed the highest agreement between the intended naturalness category and naïve observer ratings, providing the strongest perceptual exemplars of low and high naturalness. Therefore, the animations used in the present study represent the most congruent examples identified during the *Chapter 4* validation rather than a representative sample of all animations that could be generated by the reverse correlation models. Together, these materials enabled a controlled investigation of the perceptual effects of idle motion naturalness while holding visual appearance constant, thereby isolating motion as the primary experimental manipulation.

5.2.3 Procedure

The study employed a fully repeated-measures design with one within-subjects factor: Motion Condition (Still, Low Naturalness, High Naturalness). Each participant viewed all six characters in each motion condition, resulting in 18 experimental trials in total (6 characters per each of 3 motion conditions). This design allowed direct comparison of perceptual responses to different motion qualities while holding character identity constant within participants.

Upon accessing the online questionnaire, participants first reviewed the study information sheet and provided informed consent before proceeding. They were then presented with task instructions explaining that they would view a series of short video clips depicting digital characters and provide ratings after each clip. Following the instructions, participants completed three practice trials intended to familiarise them with the trial structure and rating scales. These practice stimuli featured characters not used in the experimental trials and included one example from each motion condition (still, low naturalness, high naturalness).

The animations used during practice were distinct from those included in the main experiment, ensuring that neither characters nor animation parameter sets were repeated between practice and experimental trials. Data from the practice trials were not included in the final analysis. After the practice phase, the instructions were briefly restated to remind participants of the task requirements before the experimental trials commenced.

All 18 experimental trials were presented within a single block. Trial order was fully randomised individually for each participant to minimise potential order effects and prevent systematic clustering of motion conditions. Each trial appeared on a separate questionnaire page. Participants viewed a single six-second video clip and were then required to complete all rating scales before advancing to the next trial. It was not possible to skip items, ensuring complete

data for each stimulus. The entire questionnaire took approximately 15 minutes to complete.

Participants were informed that the amount of character movement might vary across clips and that, in some cases, movement could be very subtle or the character might appear almost still. This variation was described as intentional. They were instructed to base their judgements on their overall impression of each clip rather than on the amount of movement alone, to rate each stimulus independently even if a character appeared more than once, and to use the full range of the scales. Participants were also reminded that there were no right or wrong answers and that the study was interested in their immediate impressions.

Perceptual responses were measured using validated semantic differential scales widely employed in uncanny valley research (Ho & MacDorman, 2010, 2017; Diel et al., 2021), consistent with the benchmarking methodology described in *Chapter 3, Study 1a*. Each item was rated on a 7-point bipolar scale anchored by opposing adjectives. To minimise participant fatigue given the number of stimuli, a concise subset of items capturing the core dimensions of the original indices was selected.

Human-likeness was assessed using three items (Inanimate–Living, Synthetic–Real, Human-made–Humanlike). Eeriness was measured using three items (Dull–Freaky, Predictable–Eerie, Plain–Weird). Two additional single-item scales assessed Familiarity (Strange–Familiar) and Realism (Not realistic–Very realistic), reflecting overall perceptual believability and social comfort. For each participant, mean scores were computed per dimension (human-likeness, eeriness, familiarity, and realism), with higher scores indicating greater perceived human-likeness, eeriness, familiarity, or realism, respectively. Full details of scale validation and prior usage are provided in *Chapter 3*.

Upon completion of the experimental trials, participants were given the option to enter a £50 Amazon prize draw as compensation for their time. Those who wished to participate provided an email address via a separate form that was not linked to their questionnaire responses. Email addresses were stored independently and used solely for the purpose of administering the prize draw, thereby preserving response anonymity.

This procedure enabled a controlled examination of how motion naturalness influences key uncanny valley dimensions while holding visual appearance constant across conditions.

5.2.4 Data Analysis

The primary aim was to examine whether motion naturalness influences core uncanny valley dimensions while appearance was held constant. For each participant, mean scores were computed for each perceptual dimension (Humanness, Eeriness, Familiarity, and Realism) with higher

scores indicating greater perceived humanness, eeriness, familiarity, and realism, respectively. To obtain condition-level values, ratings were averaged across all six avatars within each condition (still, low naturalness, and high naturalness). This method produced one mean score per participant for each of the three conditions.

All statistical analyses were conducted in Python (v3.11; Van Rossum and Drake Jr, 1995) using the pandas, NumPy, and SciPy stats packages. A set of one-way repeated measures ANOVAs was carried out to compare the ratings between the three motion conditions (still, low naturalness, and high naturalness) across the four perceptual dimensions (Humanness, Eeriness, Familiarity, and Realism). The significance threshold was set at $\alpha = .05$.

5.3 Results

To examine whether idle motion naturalness influenced perceptual evaluations, a series of one-way repeated-measures analyses of variance (ANOVAs) were conducted separately for human-likeness, eeriness, familiarity, and realism. Motion condition (Still, Low Naturalness, High Naturalness) was entered as a within-subjects factor. Where the assumption of sphericity was violated, Greenhouse-Geisser or Huynh-Feldt corrections were applied accordingly. Significant main effects were followed by Bonferroni-corrected pairwise comparisons. For all results, see *Figure 5.2*.

5.3.1 Human-likeness

Assumption checks indicated that the normality assumption was met for all motion conditions (Still: $W = .97, p = .46$; Low Naturalness: $W = .97, p = .47$; High Naturalness: $W = .96, p = .15$). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 19.19, p < .001$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .70$).

The repeated-measures ANOVA revealed a significant main effect of motion condition on human-likeness ratings, $F(1.40, 48.91) = 6.42, p = .008, \eta^2g = .066$, indicating that perceived human-likeness differed across motion conditions.

Bonferroni-corrected post hoc comparisons showed that characters displaying High Naturalness motion ($M = 4.28, SD = 1.36$) were rated as significantly more humanlike than those displaying Low Naturalness motion ($M = 3.52, SD = 1.19$), $t(35) = 5.62, p < .001, g = 0.59$. The difference between High Naturalness motion ($M = 4.28, SD = 1.36$) and the Still condition ($M = 3.62, SD = 1.28$) approached significance, $t(35) = 2.48, p = .054, g = 0.49$. There was

no significant difference between the Low Naturalness condition ($M = 3.52$, $SD = 1.19$) and the Still condition ($M = 3.62$, $SD = 1.28$), $t(35) = -0.36$, $p = 1.00$, $g = -0.08$.

5.3.2 Eeriness

Assumption checks indicated that normality was largely met. Shapiro-Wilk tests were non-significant for the Still ($W = .98$, $p = .86$) and Low Naturalness ($W = .96$, $p = .58$) conditions, whereas the High Naturalness condition showed a deviation from normality ($W = .93$, $p = .02$). Examination of difference scores showed that the High-Still contrast deviated from normality ($W = .93$, $p = .03$), while the remaining contrasts did not (High-Low: $W = .95$, $p = .14$; Low-Still: $W = .97$, $p = .44$). Mauchly's test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 19.01$, $p < .001$. Therefore, degrees of freedom were corrected using Greenhouse-Geisser estimates of sphericity ($\epsilon = .70$).

The repeated-measures ANOVA revealed a significant main effect of motion condition on eeriness ratings, $F(1.40, 49.01) = 25.21$, $p < .001$, $\eta^2g = .189$, indicating that perceived eeriness differed across motion conditions. As a non-parametric robustness check, a Friedman test also revealed a significant effect of motion condition on eeriness, $\chi^2(2) = 38.11$, $p < .001$.

Bonferroni-corrected post hoc comparisons showed that characters displaying High Naturalness motion ($M = 3.78$, $SD = 0.93$) were rated as significantly less eerie than those displaying Low Naturalness motion ($M = 4.74$, $SD = 1.08$), $t(35) = -7.33$, $p < .001$, $g = -0.94$. Characters in the High Naturalness condition ($M = 3.78$, $SD = 0.93$) did not differ significantly from those in the Still condition ($M = 3.67$, $SD = 1.03$), $t(35) = 0.80$, $p = 1.00$, $g = 0.11$. Because the High-Still contrast violated normality, a Wilcoxon signed-rank test was also conducted. Consistent with the paired-samples t-test, the comparison was not significant ($W = 237$, $p = .201$). In contrast, characters displaying Low Naturalness motion ($M = 4.74$, $SD = 1.08$) were rated as significantly more eerie than those in the Still condition ($M = 3.67$, $SD = 1.03$), $t(35) = 5.03$, $p < .001$, $g = 1.01$.

5.3.3 Familiarity

Assumption checks indicated that the normality assumption was met for all motion conditions (Still: $W = .97$, $p = .46$; Low Naturalness: $W = .98$, $p = .58$; High Naturalness: $W = .95$, $p = .08$). Mauchly's test indicated that the assumption of sphericity was not violated, $\chi^2(2) = 3.87$, $p = .14$.

The repeated-measures ANOVA revealed a significant main effect of motion condition on familiarity ratings, $F(2, 70) = 16.27$, $p < .001$, $\eta^2g = .128$, indicating that perceived familiarity

differed across motion conditions.

Bonferroni-corrected post hoc comparisons showed that characters displaying High Naturalness motion ($M = 4.28$, $SD = 1.19$) were rated as significantly more familiar than those displaying Low Naturalness motion ($M = 3.26$, $SD = 1.07$), $t(35) = 6.57$, $p < .001$, $g = 0.89$. There was no significant difference between the High Naturalness condition ($M = 4.28$, $SD = 1.19$) and the Still condition ($M = 4.06$, $SD = 1.20$), $t(35) = 1.11$, $p = .82$, $g = 0.18$. Finally, characters displaying Low Naturalness motion ($M = 3.26$, $SD = 1.07$) were rated as significantly less familiar than those in the Still condition ($M = 4.06$, $SD = 1.20$), $t(35) = -3.85$, $p = .002$, $g = -0.69$.

5.3.4 Realism

Assumption checks indicated that the normality assumption was met for all motion conditions (Still: $W = .99$, $p = .90$; Low Naturalness: $W = .98$, $p = .71$; High Naturalness: $W = .98$, $p = .67$). Mauchly's test indicated that the assumption of sphericity was violated, $\chi^2(2) = 10.70$, $p = .005$. Because the estimate of sphericity exceeded .75 ($\epsilon = .79$), Huynh-Feldt corrections were applied.

The repeated-measures ANOVA revealed a significant main effect of motion condition on realism ratings, $F(1.57, 55.12) = 9.97$, $p = .001$, $\eta^2g = .088$, indicating that perceived realism differed across motion conditions.

Bonferroni-corrected post hoc comparisons showed that characters displaying High Naturalness motion ($M = 4.53$, $SD = 1.31$) were rated as significantly more realistic than those displaying Low Naturalness motion ($M = 3.54$, $SD = 1.25$), $t(35) = 6.38$, $p < .001$, $g = 0.76$. There was no significant difference between the High Naturalness condition ($M = 4.53$, $SD = 1.31$) and the Still condition ($M = 4.15$, $SD = 1.41$), $t(35) = 1.51$, $p = .42$, $g = 0.28$. Finally, characters displaying Low Naturalness motion ($M = 3.54$, $SD = 1.25$) tended to be rated as less realistic than those in the Still condition ($M = 4.15$, $SD = 1.41$), although this difference did not reach significance, $t(35) = -2.45$, $p = .059$, $g = -0.45$.

Perceptual Ratings Across Motion Conditions

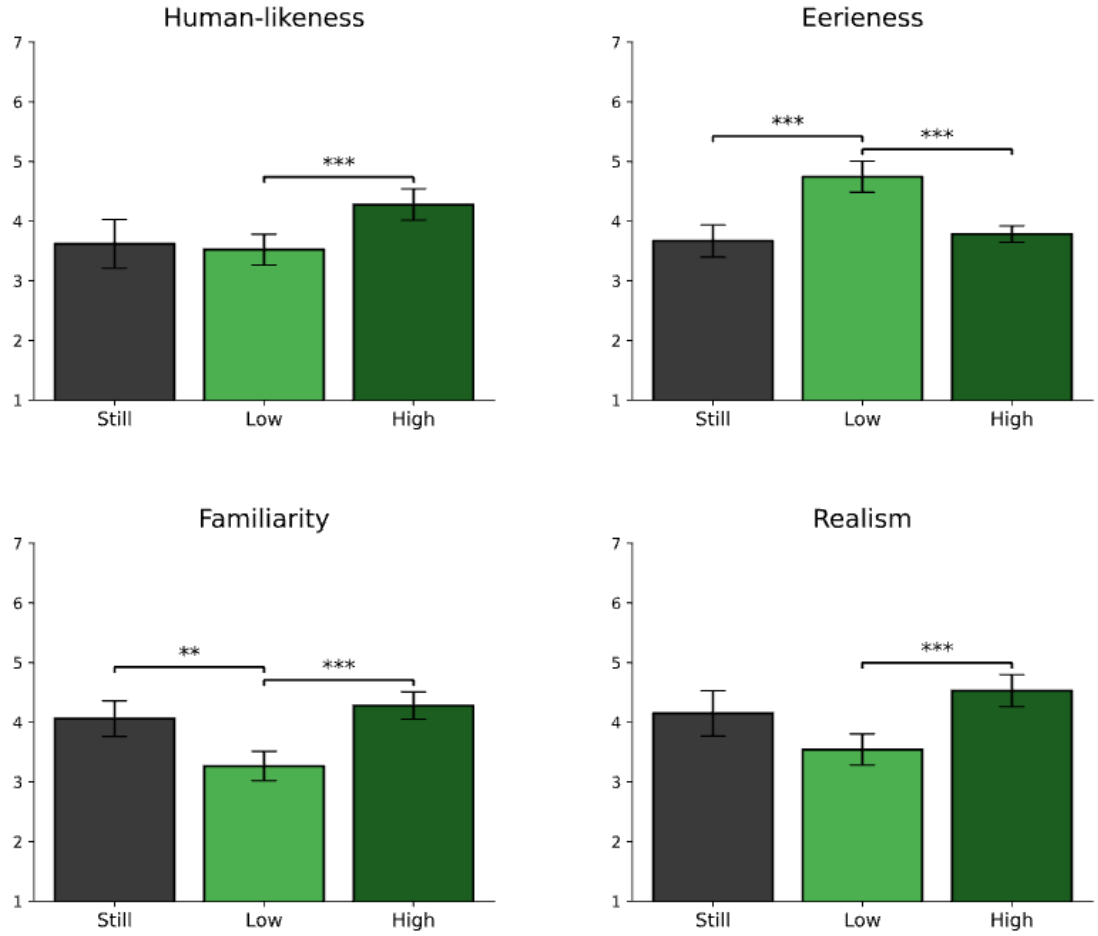


Figure 5.2: Mean ratings (1-7 scale) of human-likeness, eeriness, familiarity, and realism as a function of idle motion condition (Still, Low Naturalness, High Naturalness). Error bars represent 95% within-subject confidence intervals (Cousineau-Morey corrected). Asterisks indicate significant pairwise comparisons (Bonferroni corrected): * $p < .05$, ** $p < .01$, *** $p < .001$.

5.4 Discussion

The present study aimed to isolate motion naturalness as a causal factor in core uncanny valley dimensions while holding visual appearance constant. Building on the validation of the MetaHuman platform established in *Chapter 3* and the perceptually grounded reverse-correlation models of idle motion naturalness developed in *Chapter 4*, the present experiment conducted a controlled test of mismatch-based accounts of the uncanny valley (Di Natale et al., 2023; Kätsyri et al., 2015). By presenting the same high-fidelity characters in three conditions (still, low naturalness motion, and high naturalness motion), the study was able to separate the effects of having motion at all from the effects of how natural that motion was. Because the characters' visual

appearance remained identical across conditions, any differences in evaluation can be attributed to motion rather than rendering quality or facial realism. This is important, as previous research has often varied appearance and movement at the same time, making it difficult to determine whether uncanny responses are driven by dynamic properties, visual fidelity, or their interaction (Crookes et al., 2015; Kätsyri et al., 2015; McDonnell et al., 2012). The key finding was clear: unnatural idle motion significantly increased eeriness and reduced positive evaluations relative to high naturalness motion, and in some cases relative to stillness. In contrast, natural idle motion largely preserved baseline impressions associated with stillness. In other words, introducing biologically implausible dynamics imposed a perceptual cost, whereas introducing biologically plausible idle behaviour did not add uncanniness and showed only limited evidence of additional perceptual benefit beyond the still condition.

5.4.1 Eeriness and Dynamic Cue Mismatch

The eeriness dimension offered the most direct test of mismatch-based interpretations of the uncanny valley. As predicted, characters displaying low naturalness idle motion were rated as significantly more eerie than both still characters and those displaying high naturalness motion, whereas still and high naturalness conditions did not differ from one another. This pattern closely aligns with perceptual mismatch and predictive processing accounts (Di Natale et al., 2023; Diel & Lewis, 2022; Kätsyri et al., 2015), which propose that discomfort arises when expected biological dynamics are violated. In the present case, exaggerated or atypical idle facial movements appear to have introduced incoherent dynamic cues relative to the otherwise highly realistic MetaHuman appearance, thereby elevating eeriness. This interpretation is consistent with experimental work showing that subtle temporal distortions in facial animation increase uncanniness across both human and artificial agents (Diel et al., 2023), as well as neural evidence indicating sensitivity to mismatches between appearance and motion (Saygin et al., 2012).

Importantly, the absence of a difference between stillness and high naturalness motion suggests that biologically plausible idle dynamics do not, in themselves, reduce eeriness below a static baseline. Rather than observing a strong ‘bonus’ effect of natural motion, the pattern is consistent with a cost of mismatch. Natural idle motion maintained perceptual coherence and avoided introducing additional uncanniness, but it did not significantly improve affective evaluations beyond the level observed for still characters. This asymmetry, where incoherence produces discomfort but coherence merely preserves neutrality, provides nuanced support for mismatch-based theories and suggests that uncanny responses may be driven more by violations than by enhancements of cue consistency.

5.4.2 Motion Quality Beyond Eeriness

Across humanlikeness, familiarity, and realism, a consistent pattern emerged: characters displaying high naturalness motion were rated significantly more positively than those displaying low naturalness motion. This robust difference across all three dimensions indicates that motion quality exerts a broad influence on evaluative judgements, extending beyond affective discomfort (eeriness) to encompass more general impressions. In line with prior work demonstrating that biologically plausible motion enhances perceived acceptability and humanlikeness (Asselborn et al., 2017; Piwek et al., 2014; Treal et al., 2020, 2021), unnatural idle dynamics appear to undermine these impressions when they violate expectations of subtle, neutral facial behaviour.

However, when comparing high naturalness motion to the still baseline, the pattern was more nuanced. For familiarity and realism, no significant differences were observed between still and high naturalness conditions. For humanlikeness, no significant difference was observed between the high naturalness and still conditions. This suggests that while natural motion clearly outperforms unnatural motion, it does not necessarily produce evaluations that exceed those elicited by a static, visually coherent character. Rather than functioning as a guaranteed ‘upgrade’, natural idle motion appears to act predominantly as a stabilising cue, maintaining perceptual coherence without necessarily amplifying positive impressions.

On the other hand, comparisons between the low naturalness and still conditions highlight the cost of mismatch more clearly. Familiarity ratings were significantly lower in the low naturalness condition compared to the still condition. This finding suggests that introducing biologically implausible motion can negatively influence familiarity judgements, supporting the view that evaluative penalties may arise from dynamic incoherence rather than from the absence of motion itself. No significant difference was observed between the low naturalness and still conditions for realism.

One interpretation is that unnatural idle movements disrupt the smooth integration of appearance and motion. When the face does not move in a way that matches expectations, it becomes harder to process as a coherent human agent, which may reduce feelings of familiarity and realism. Although the present study did not directly manipulate appearance and motion independently to test cue integration, the observed effects may still be consistent with cue-integration accounts. In this view, violations of natural motion introduce dynamic irregularities that make the character harder to process as a coherent human agent. Such inconsistencies could therefore influence broader evaluative judgements, extending the cost of mismatch beyond eeriness into more general social and perceptual impressions.

5.4.3 Methodological Contribution

Beyond its theoretical implications, the present study makes an important methodological contribution to uncanny valley research. Although motion has long been proposed as central to uncanny responses (Kätsyri et al., 2015; Mori, 1970; Mori et al., 2012), empirical investigations have often relied on lower-fidelity characters, broad categorical contrasts (e.g., biological vs mechanical motion), or designs in which motion and appearance vary simultaneously. In many cases, appearance realism is the primary manipulated variable, with motion quality either loosely defined or confounded with rendering differences. Such designs make it difficult to isolate the specific contribution of motion itself.

In contrast, the current experiment was explicitly designed to isolate motion naturalness as a causal factor while holding visual appearance constant. A validated high-fidelity MetaHuman baseline (*Study 1a*; *Study 1b*) ensured that the characters' appearance did not inherently elicit strong uncanny responses. Motion was manipulated using reverse-correlated, perceptually grounded models derived directly from observer judgements (*Study 2*), rather than arbitrary distortions or broad biological vs mechanical contrasts. Importantly, three clearly defined motion conditions were implemented: no movement (still), biologically plausible (natural), and biologically implausible (unnatural) motion. The inclusion of a still condition allowed dissociation between motion presence (still vs animated) and motion quality (low vs high naturalness).

Taken together, these features represent one of the most tightly controlled operationalisations of motion within uncanny valley research to date. By holding appearance constant and independently manipulating motion quality, the design allowed direct comparison between the absence of movement, the addition of coherent natural movement, and the addition of unnatural movement. Comparing still and natural motion tests whether adding a biologically plausible dynamic cue affects evaluations relative to a static baseline. Comparing still and unnatural motion examines the effect of introducing an mismatching/conflicting dynamic cue. Additionally, comparing natural and unnatural motion isolates the impact of motion quality itself, as both conditions involve movement and differ only in how natural it is.

Since rendering, texture, lighting, and facial structure were identical across all conditions, any observed differences in eeriness, humanness, familiarity, or realism can be attributed specifically to motion naturalness rather than to graphical artefacts or changes in visual realism. This level of control enabled a clean causal test of motion-based mismatch independent of appearance effects.

It should also be noted that the animations used in the present study were selected from the most congruent exemplars identified during the validation phase reported in *Chapter 4*. Thus, the observed effect sizes should be interpreted as illustrating what the reverse correlation mod-

els can achieve under optimal conditions, rather than representing the effects expected across all generated animations. As discussed in *Chapter 4*, the reverse correlation analysis treated facial AUs independently and did not model interactions between co-occurring facial movements. It is therefore possible that the highly congruent animations used in the present study reflect AU combinations that happened to produce particularly coherent idle facial behaviour. Future work should investigate how combinations of co-occurring AUs influence perceived naturalness, with the aim of developing models that more consistently generate highly congruent natural and unnatural animations.

5.5 Conclusion

The present study demonstrates that motion naturalness plays a critical role in uncanny valley evaluations when realistic appearance is held constant. Unnatural motion consistently increased eeriness and reduced humanness, familiarity, and realism relative to natural motion. In contrast, natural motion largely preserved baseline impressions associated with still characters and showed only limited evidence of additional benefit. Overall, the pattern was asymmetric: negative reactions emerged when motion violated expectations, whereas natural motion largely preserved baseline impressions rather than substantially enhancing them.

These findings provide direct support for mismatch-based accounts of the uncanny valley (Di Natale et al., 2023; Diel & Lewis, 2022; Kätsyri et al., 2015). The results indicate that evaluative penalties arise when dynamic cues conflict with a realistic appearance, whereas coherent motion does not dramatically elevate impressions beyond baseline. In this context, uncanny responses appear to be driven more by violations of expected dynamics than by the mere presence of movement.

Methodologically, by combining a validated high-fidelity MetaHuman baseline with perceptually grounded, reverse-correlated motion manipulations, the study provides a clean causal test of motion-based mismatch independent of appearance effects. From a practical perspective, the results suggest that avoiding unnatural or implausible motion may be more important than adding increasingly complex behaviour. Natural idle motion does not dramatically improve impressions, but unnatural motion clearly damages them. In realistic digital humans, coherent motion is perceptually safe, whereas incoherent motion produces reliable negative reactions. It is also worth noting that complete stillness is largely an experimental simplification. In natural social interactions, faces exhibit continuous subtle movements, and future work should therefore continue refining how different forms of motion contribute to perceived aliveness and social engagement.

Chapter 6

General Discussion

The present thesis set out to understand how digital humans come to feel “alive” rather than uncanny, with a focus on the perceptual role of subtle, non-expressive behaviour. Across the empirical chapters, the work established a pipeline that begins with a controlled, high-fidelity stimulus platform, then derives perceptually grounded models of natural idle facial motion, and finally, tests whether manipulating motion naturalness causally changes core uncanny valley evaluations. In this thesis, these evaluations are captured through ratings of humanness, eeriness, realism, and familiarity, which together reflect how human-like, comfortable, and believable digital characters appear to observers. The discussion begins by summarising the contribution of each chapter to that pipeline, and then extends out to the theoretical, methodological, and applied implications of the thesis for research on social perception, reverse correlation and the uncanny valley.

6.1 Contributions of Each Chapter

Overall, the thesis emphasises one central claim. For realistic digital humans, there is a significant and consistent perceptual cost of unnatural motion, while the perceptual benefit of adding natural motion is smaller and more conditional. This asymmetry is important because it shifts the emphasis away from realism as a simple route to positive impressions, and towards cue coherence as a requirement for avoiding negative impressions. Each chapter contributes a distinct piece of evidence to support this claim.

Chapter 3 established the perceptual baseline by directly comparing matched avatar identities across the MetaHuman and GFG frameworks, such that each character was represented in both systems and differed primarily in graphical fidelity. This comparison allowed perceptual differences to be attributed more specifically to rendering and appearance quality, rather than to

variation in character design, and provided a validated reference point for how the avatars were perceived in the absence of motion manipulations. Within this controlled comparison, MetaHuman avatars were rated as more humanlike and more realistic whilst also being perceived as less eerie than their GFG counterparts. This is important because uncanny valley questions are difficult to interpret when the stimulus platform is limited by visual artefacts that can drive discomfort.

Earlier avatar systems used in social perception and reverse correlation research have strong experimental advantages; yet, as established in the general introduction, their graphical fidelity reflects the standards of their time (Jack & Schyns, 2017; Todorov et al., 2011). This includes limitations in facial shape accuracy, skin rendering, and dynamic behaviour, all of which are known to influence how faces are perceived. For instance, computer-generated faces can engage face processing mechanisms differently from real faces, showing reduced recognition performance and weaker markers of perceptual expertise (Crookes et al., 2015). At the same time, research on the uncanny valley shows that even subtle inconsistencies in appearance, such as mismatches between realistic textures and atypical facial proportions or feature configurations, can increase eeriness (Chattopadhyay & MacDorman, 2016; MacDorman et al., 2009; Seyama & Nagayama, 2007). Appearance cues are also not limited to the face alone, as factors such as voice, body movement, and cross-modal coherence can independently contribute to perceived human-likeness or uncanniness when they are inconsistent (Mitchell et al., 2011; Kätsyri et al., 2015).

Such appearance-related factors may therefore create ambiguity about whether observers are responding to behavioural cues or to low-level rendering artefacts, especially when research aims to interpret outcomes in terms of mismatch between appearance and motion (Kätsyri et al., 2015; McDonnell et al., 2012). By benchmarking MetaHumans against the GFG platform using validated measures of uncanny valley-related impressions, the thesis showed that high-fidelity appearance can improve core evaluations such as humanness and realism without necessarily increasing eeriness, in line with work suggesting that state-of-the-art photorealism can avoid the valley when visual cues remain coherent (Bae et al., 2025; Di Natale et al., 2023; Higgins et al., 2021). This chapter therefore contributed a controlled appearance foundation in the form of MetaHuman avatars that reduces an important confound in later motion manipulations.

Chapter 4 developed a perceptually grounded model of idle facial motion naturalness. Most reverse correlation studies in face perception have focused on socially interpretable targets such as emotions, traits, and mental states, using either noise-based classification images or parametric generative spaces (Dotsch & Todorov, 2012; Jack & Schyns, 2017; Todorov et al., 2025). Dynamic reverse correlation has successfully recovered the facial action unit patterns and temporal dynamics underlying emotion communication (Jack et al., 2014; Yu et al., 2012). The present thesis extended this approach to a more subtle target. Idle motion does not convey a

clear communicative message, yet prior work in robotics and avatar perception suggests that such “background” movement contributes to animacy and believability (Asselborn et al., 2017; Treal et al., 2020, 2021), and recent work has argued that idle behaviour is understudied and difficult to author or capture convincingly (Landa et al., 2025). *Chapter 4* contributed evidence that idle facial motion contains structured information that observers use when judging its naturalness. Low naturalness was associated with larger lower face movements such as jaw-dropping and lip-parting actions that seem unusual in neutral idling, while high naturalness relied more on constrained, localised actions such as subtle lip-pressing and brow activity. These patterns suggest that perceived naturalness depends less on the overall amount of movement and more on whether the type and intensity of movement are likely for a neutral idle context. The validation phase added a key element that goes beyond correlational description. Naïve observers distinguished low and high naturalness animations at above chance levels, suggesting that the models were generative in a perceptually meaningful sense, rather than simply summarising the characteristics of the original dataset.

Chapter 5 tested whether motion naturalness affects core uncanny valley dimensions when appearance is held constant. Uncanny valley research has often relied on broad contrasts, including categorical manipulations of biological versus mechanical motion (Saygin et al., 2012), or designs where appearance and motion vary together, making it difficult to isolate the contribution of dynamics (Di Natale et al., 2023; Kätsyri et al., 2015). *Chapter 5* contributed a clean causal test by combining the validated appearance baseline from *Chapter 3* with the perceptually grounded motion manipulation from *Chapter 4*. The results reflected a consistent pattern across outcomes. Low naturalness motion consistently increased eeriness and reduced humaneness, familiarity, and realism relative to high naturalness motion. The still condition provided a further insight. Stillness was generally aligned more closely with high naturalness than low naturalness on eeriness, familiarity, and realism. This suggests that the cost of adding mismatching motion can exceed the cost of not introducing motion at all, which matches theoretical expectations from mismatch accounts (Kätsyri et al., 2015). High naturalness motion mostly preserved baseline impressions rather than producing large improvements over stillness. No significant differences were observed between the high naturalness and still conditions across the measured dimensions. The overall pattern therefore supports an asymmetric interpretation in which incoherent dynamics produce a penalty, while coherent dynamics keep evaluations stable. At the same time, this raises an important question about what is required for a character to appear truly “alive” rather than simply not uncanny. While stillness may avoid perceptual conflict, it may also limit impressions of animacy and social presence, suggesting that naturalness alone may not be sufficient, and that some degree of subtle, biologically plausible movement is likely necessary to convey aliveness.

Taken together, these chapter-level contributions form a robust and integrated pipeline. The

perceptual advantage of a high-fidelity MetaHuman platform over earlier avatar systems used in social psychophysics was empirically validated rather than assumed, motion naturalness was derived from human perception rather than more arbitrary manipulations, and motion effects were tested causally using a design that separates motion presence from motion quality.

6.2 What the Thesis Changes About Motion and the Uncanny Valley

A broad implication of the thesis is the reinforcement of the idea that the uncanny valley is best approached as a problem of perceptual coherence rather than a simple function of realism, while also suggesting that the construct itself may benefit from a broader definition; one that encompasses not only eeriness and humanlikeness but a broader range of evaluative dimensions including familiarity and realism, and potentially others yet to be established. Mori's original proposal suggested that affinity increases with human likeness until a sharp drop occurs for near-human entities (Mori, 1970; Mori et al., 2012). Later empirical work revealed mixed support for a single realism driven curve, and reviews emphasised that multiple mechanisms may contribute to the eeriness perceptions (Di Natale et al., 2023; Kätsyri et al., 2015). The present thesis adds evidence that, for high fidelity digital humans, motion naturalness plays a direct causal role in where a character sits on the key evaluative dimensions. Chapter 5 shows that adding unnatural idle motion can increase eeriness when appearance is highly realistic. This aligns with work showing that subtle temporal disruptions to facial expressions increase uncanniness across humans, androids, and virtual faces (Diel et al., 2023), and it is consistent with predictive processing interpretations that emphasise violations of expected dynamics (Diel & Lewis, 2022). It also fits neural evidence that a humanlike appearance paired with “non-biological” movement elicits distinct activity in the action perception brain networks (Saygin et al., 2012). The present contribution demonstrated this logic with a more controlled and perceptually derived manipulation, rather than a broad biological-mechanical contrast of motion.

The still condition is also theoretically informative. Stillness removes dynamic cues to animacy, yet it does not necessarily create the kind of conflict that mismatch accounts predict. In *Chapter 5*, low naturalness motion produced more negative evaluations than stillness on several uncanny valley measures, particularly eeriness and familiarity. This suggests that discomfort is driven strongly by the presence of conflicting dynamic information, rather than by the absence of motion cues. This distinction matters because many classic studies have used static faces as stimuli by default (Kätsyri et al., 2015). Under a mismatch view, static faces do not create a conflict between channels because motion cues are absent altogether, whereas unnatural motion can actively contradict appearance-based expectations by introducing a new cue in the form of

motion. At the same time, work in human-robot interaction has shown that introducing subtle idle movements during inactive moments can improve interaction quality (Asselborn et al., 2017). One interpretation of their findings, consistent with the present results, is that complete stillness during expected behavioural pauses may itself function as an incoherent cue in certain contexts, particularly when an agent is otherwise framed as socially responsive as was the case in Asselborn et al.’s study. Although this remains speculative, it suggests that the perceptual consequences of stillness may depend on the expectations established by the interaction context. The thesis therefore supports a view in which motion amplifies uncanny responses only when it violates expected patterns of human movement in the given context. In this sense, the effect reflects a perceptual penalty for expectation violation rather than a direct consequence of increased realism or unnatural movement.

The asymmetric pattern observed in *Chapter 5* also helps interpret the role of motion in positive uncanny valley related impressions. A common intuition is that adding motion should make the characters appear more alive, which in turn should increase humanness and reduce eeriness. The thesis suggests a more constrained story. Coherent idle motion appears perceptually safe, and it can support human like impressions, yet its effect may be modest when the baseline appearance is already high fidelity and socially acceptable. This interpretation aligns with the idea that some cues mainly prevent negative reactions rather than producing strong positive gains. It is also compatible with findings that realism does not inevitably increase discomfort, and that high fidelity characters can be rated positively when cues are consistent (Bae et al., 2025; Higgins et al., 2021). In practice, this implies that designers and researchers may see greater benefits from removing implausible motion than from adding more complex motion in certain contexts.

6.3 Extending Reverse Correlation to Idle Facial Motion

A second broad implication is that reverse correlation can be used to model subtle behavioural realism, not just expressive communication. Reverse correlation has a long history in psychophysics as a way to infer the stimulus features that drive impressions (Ahumada & Lovell, 1971). In social perception, it has been used to visualise mental representations of faces and traits, often through noise-based methods or controlled generative spaces (Brinkman et al., 2017; Dotsch & Todorov, 2012; Gosselin & Schyns, 2003). Dynamic variants have also linked temporal facial action unit patterns to emotion perception and intensity judgements (Jack et al., 2014; Yu et al., 2012). These methods typically target stimuli with clear social meaning, which makes the link between movement and judgement relatively direct.

Idle facial motion is different. It is subtle, non-expressive, and often treated as a background

in animation practice, yet it is likely to contribute to a basic impression that a face is alive (Bates, 1994; Kocoń, 2013). *Chapter 4* demonstrates that observers can still provide consistent information about such subtle, non-expressive motion when asked to judge naturalness, and that the resulting models can generate new stimuli with predictable perceptual outcomes. This extends the scope of reverse correlation from modelling communicative signals to modelling background dynamics that function as cues to animacy. It also complements work suggesting that idle behaviour is difficult to capture convincingly and that manually-animated idles are perceptually distinct from motion-captured ones (Landa et al., 2025). Instead of treating naturalness as a property that must be imported from motion capture, the thesis operationalises natural idle motion as a perceptual judgement that can be modelled and subsequently synthesised.

The mid-range naturalness ambiguity observed in *Chapter 4* is also theoretically meaningful. The validation data suggest that low and high naturalness are clearer categories for observers, while mid naturalness is less stable. This supports an interpretation of naturalness as a graded perceptual dimension rather than a sharply defined category. Under predictive processing views, extremes may correspond to strong confirmation or strong violation of expectations, while intermediate cases contain mixed evidence, producing variability in category placement across observers and trials (Diel & Lewis, 2022). At the same time, certain stimuli were correctly judged as mid naturalness with relatively high consistency across participants, suggesting that intermediate impressions are not purely noisy or unstable. One possibility is that specific combinations of AU activations interact in systematic ways to produce perceptually coherent mid-level naturalness signals. Future research could therefore examine how particular AU patterns and their temporal relationships combine to generate low, mid, and high naturalness impressions, with the aim of refining synthesis models and improving classification accuracy across the full idle motion naturalness spectrum.

6.4 Methodological Contributions

The methodological contribution of this thesis lies in its development towards a tightly controlled and clearly defined test of motion naturalness. Each chapter added a cumulative element that culminated in the final study of the thesis. *Chapter 3* showed that a high-fidelity MetaHuman platform could serve as a perceptually validated appearance baseline. *Chapter 4* used that platform to derive generative models of natural idle motion directly from observer judgements, allowing motion naturalness to be precisely manipulated. *Chapter 5* brought these elements together by manipulating motion naturalness within that validated appearance framework to test its direct impact on uncanny valley dimensions. This cumulative structure strengthens internal validity by ensuring that motion effects are evaluated against a validated appearance foundation and that behavioural outcomes are tested using stimuli whose properties were derived from

human perception rather than loosely defined or arbitrary motion manipulations.

The use of MetaHumans is particularly impactful here. Dynamic and generative model reverse correlation has historically relied on platforms, such as FaceGen and GFG (Oosterhof & Todorov, 2009; Yu et al., 2012), that support high experimental control but limited graphical fidelity. By implementing the procedure in Unreal Engine with industry-grade state-of-the-art MetaHuman characters, the thesis demonstrates that perception driven modelling can be combined with contemporary graphics pipelines whilst maintaining the ability to scale stimulus generation and retain precise experimental control. This is valuable for three reasons. Firstly, it means that the research is conducted using the same character systems and animation pipelines that are widely used in contemporary video games and digital media. This increases the practical relevance of reverse correlation methods and opens clearer pathways for collaboration between academic research and industry practice. Secondly, it allows theoretical questions about realism, mismatch, and animacy to be tested using stimuli that better match what observers encounter in real life and modern digital environments. Thirdly, MetaHuman is an open and freely available platform, which lowers barriers to entry for researchers and may accelerate methodological development in the field. In contrast to paid and proprietary systems such as FaceGen and the GFG respectively, the accessibility of MetaHuman increases the potential for replication, extension, and wider adoption across laboratories worldwide.

The design of *Chapter 5* is also methodologically distinctive. It introduces a still condition alongside low and high naturalness motion, which enables separation of motion presence effects from motion quality effects. Many studies compare one moving condition to another, which can conflate changes in amplitude, timing, and perceived plausibility. Others compare moving to still, which can conflate presence with quality. By including all three, *Chapter 5* makes it possible to ask a more precise question. Does motion itself alter impressions? Does degraded motion harm impressions? The results suggest that degraded motion produces the clearest effect. This supports a view that the crucial variable in uncanny responses is coherence of dynamic cues, which is a testable claim that the thesis operationalises with clarity.

6.5 Practical Implications for Digital Human Creation

The practical implication of the thesis is that avoiding implausible idle motion is a high priority. In production pipelines, idle behaviour is often, reused across characters and generated procedurally for scalability. These practices can be effective, yet they can also introduce motion that is exaggerated, repetitive, or context inappropriate. The present work suggests that natural motion can carry a strong perceptual penalty when the character looks highly realistic, because it creates a mismatch between appearance and dynamics. This aligns with evidence that motion

anomalies interact with render style and can be judged more negatively on realistically rendered characters compared to stylised ones (McDonnell et al., 2012). It also aligns with findings that subtle background movement can increase believability and social response (Treal et al., 2020, 2021), and that idle motion can increase perceived friendliness and human likeness in robots (Asselborn et al., 2017). The thesis adds that, when motion is implausible, it can push evaluations in the negative direction across several dimensions, including eeriness and familiarity, highlighting that the key design variable is not the presence of motion itself but its coherence with the character’s appearance and expected movement patterns.

The reverse correlation framework also suggests a solution for scalable content creation. Rather than relying on intuition, or difficult to record motion capture, for natural idling, the perceptual models derived in *Chapter 4* provide a parameter space grounded in observer judgements. In principle, this data-driven approach could feed into animation libraries, procedural generators, and quality control tools, especially for background characters in large-scale projects where bespoke motion capture is impractical. The point is not that perception driven synthesis replaces motion capture or skilled animation. The point is that it can provide a scalable method for generating subtle, plausible behaviour that avoids the most perceptually costly mismatches.

More broadly, implementing reverse correlation within the MetaHuman ecosystem demonstrates that perception driven modelling can operate directly within the same character systems and animation pipelines used in contemporary video games and digital media. This shared platform reduces the gap between laboratory research and production environments. Because the stimuli are generated within an industry standard framework, perceptual findings can be translated more directly into existing animation workflows, allowing empirical research to inform design decisions more practically. Such alignment creates opportunities for closer collaboration between academia and industry, where advances in perceptual modelling can contribute to the digital character production workflows in the industry.

6.6 Limitations

Several limitations should be considered when interpreting the findings. Firstly, the modelling of head and eye motion in *Chapter 4* showed weaker relationships with naturalness ratings in the main dataset where they appeared along with facial motion, which led to an additional dataset without facial AUs. This suggests that salience and cue weighting may shift depending on what other cues are present. In natural perception, observers integrate information across the face, head, and body. When one channel contains large changes, it can dominate judgements (Aviezer et al., 2012). This indicates that reverse correlation findings may depend on how multiple behavioural channels are combined and how salient each channel is. Future work can address this

by explicitly manipulating cue/feature salience, or by modelling naturalness using multivariate approaches that capture interactions between channels.

Secondly, the data for *Chapter 5* was collected online using brief clips embedded in a questionnaire. Online data collection provides scalability and diversity, yet it reduces control over viewing conditions such as screen size or viewing environment. These factors may affect sensitivity to subtle motion. The fact that robust differences emerged despite these constraints strengthens the claim that motion naturalness has meaningful perceptual consequences. Future work could test whether effects uphold or become amplified under more controlled viewing conditions.

Thirdly, statistical power should also be considered when interpreting the present findings. Although several robust effects were observed across the studies, some comparisons did not reach statistical significance despite yielding small-to-medium or medium effect sizes. Consequently, it remains possible that some null findings reflect limited statistical power rather than the complete absence of an underlying effect. Future studies with larger samples would provide greater power to detect more subtle differences and help determine whether these null findings reflect genuine perceptual equivalence or smaller effects that were not detectable in the present experiments.

Finally, the work focused on neutral, non-communicative idle motion. Many of the strongest uncanny effects in real world contexts may involve speech, emotional expressions, and interactive timing. The present thesis isolates idle motion because it is a foundational cue to “aliveness” and because it is typically present even when a character is not actively communicating. In applied systems, this kind of idle behaviour can be treated as a base that other animations are layered on top of, rather than as an isolated clip. For example, a virtual receptionist or game companion may spend several seconds waiting for the user to speak or decide what to do. If the character is perfectly still during that waiting period and then abruptly switches into a smile or a speaking animation, the transition can feel mechanical or poorly timed, even if the expressive animation itself is well made. By contrast, an avatar that maintains subtle, coherent idle motion while waiting may preserve continuity, so that a later shift into a smile, eye contact, or speech onset reads as a natural change in state rather than a hard “mode switch”. The results therefore provide a basis for future work that examines how idle motion interacts with expressive behaviour, turn taking, and social responsiveness, which are known to shape impressions of artificial agents (Krämer et al., 2018).

6.7 Future Directions

Thus, the thesis opens several promising directions. One direction is to extend perceptual modelling beyond the face to full body behaviour. Postural oscillations have been shown to increase believability and affective response to avatar expressions (Treal et al., 2020, 2021), and body motion can capture attention independently of appearance (Williams et al., 2019). With the full body MetaHuman rigs, reverse correlation could be extended to body idling and posture, allowing a more complete model of aliveness cues across channels.

A second direction is to model individual differences in sensitivity to mismatch. Reverse correlation has been used to reveal variability in mental representations linked to beliefs and attitudes in other domains (Brown-Iannuzzi et al., 2017; Young et al., 2014). It is plausible that individuals differ in how strongly they weight motion cues when judging eeriness, or in what kinds of motion they treat as implausible. This notion has also been supported by mixed perceptions of the different characters in the qualitative portion of Chapter 3. Identifying such differences could help explain variability in uncanny valley effects across studies (Kätsyri et al., 2015).

A third direction would be to systematically operationalise perceived naturalness itself. Throughout the present thesis, naturalness was intentionally treated as an intuitive perceptual judgement because there is currently no widely accepted operational definition of what constitutes natural idle facial behaviour (Lei et al., 2021; Pollick et al., 2005; Voas et al., 2023). Future research could seek to identify the specific perceptual dimensions that underpin these judgements, for example by examining the contribution of movement amplitude, temporal smoothness, biological plausibility, and the coordination between facial, head, and eye movements. It will also be important to determine how these judgements vary across different behavioural contexts, as idle behaviour that appears natural while listening, waiting, or thinking may differ from behaviour perceived as natural during social interaction or other situations. Establishing a more systematic operational framework for perceived naturalness would improve comparability across studies and provide a stronger theoretical basis for future perceptual modelling of digital human behaviour.

A fourth direction is to refine mid range modelling. Although medium naturalness was generally less stable than the low and high categories, some stimuli were judged as mid naturalness with high consistency across participants. This suggests that certain combinations of action units may reliably produce intermediate impressions, while others generate more ambiguous responses. Future work could therefore examine how specific AU activation patterns interact to produce low, medium, and high naturalness, and use this knowledge to synthesise animations that elicit more precise and consistent perceptions.

A fifth direction concerns the measurement framework used to capture uncanny valley responses, and how the evaluative questions asked shape what kind of uncanny valley is observed. The present thesis assessed motion naturalness across four dimensions of eeriness, humanness, familiarity, and realism, which all showed significant main effects of motion condition in *Study 3*, but with notably different patterns of sensitivity. This raises questions about whether existing measures are sufficiently discriminable, and whether additional dimensions could increase sensitivity to specific aspects of the uncanny valley.

Eeriness showed the largest and most consistent effect, with the clearest separation between all three conditions and the highest effect size. Familiarity showed a significant cost of unnatural motion relative to stillness, whereas no significant difference was observed for realism when compared to the still condition. Likewise, no significant difference was observed between the high naturalness and still conditions for humanness. These differences suggest that the four dimensions are not capturing the same underlying response, and that they may be differentially sensitive to motion-based manipulations. Future work could therefore examine which dimensions are most sensitive to specific types of motion perturbation, and whether some, such as realism and familiarity, are better understood as partially overlapping indicators of perceptual coherence or whether they can be more cleanly dissociated through targeted experimental designs. Diel et al.'s (2021) meta-analysis supports this kind of interrogation, identifying 39 different affect measures across the uncanny valley literature and showing that effect sizes vary substantially depending on the measure used, with animacy and experience-related dimensions showing particularly large effects, suggesting these may offer greater sensitivity to subtle dynamic cues of the kind examined here. This points to a broader conceptual direction motivated by converging evidence from Chapter 3.

The quantitative findings of *Study 1b* showed that MetaHuman avatars were rated significantly higher than GFG avatars on emotional capacity, person-likeness, and moral concern, dimensions that go beyond standard uncanny valley measures and reflect higher-level socio-cognitive attributions. Importantly, these effects emerged alongside, but were not reducible to, the humanness and eeriness differences observed in *Study 1a*, suggesting that observers were making evaluative judgements that existing uncanny valley indices do not fully capture. This was further reinforced by the qualitative findings of *Study 1b*, in which participants described a sense of inner life, depth, and social presence in the avatars that similarly did not map straightforwardly onto humanness or eeriness alone. While MetaHumans were consistently rated as more humanlike in the quantitative data, qualitative responses revealed considerable variation in how participants interpreted this realism socially, with some describing MetaHumans as emotionally distant or hollow despite their greater visual fidelity, and others attributing greater emotional presence to the less realistic GFG avatars.

Together, these quantitative and qualitative findings suggest that what observers are actually

responding to in realistic digital humans may be something closer to perceived social aliveness, that is, whether a character feels internally present, responsive, and capable of engaging socially, rather than simply whether it looks human or feels eerie. Asking whether a character is socially alive, rather than whether it is humanlike or uncanny, would constitute a meaningfully different question, and one that might reveal a different kind of uncanny valley. A character could be high in humanness and low in eeriness while still feeling hollow or socially absent, a pattern that standard indices would not capture but that socially focused measures might. Developing and validating measures of perceived social aliveness or social responsiveness, informed by both the quantitative socio-cognitive dimensions and qualitative themes identified in *Chapter 3*, therefore represents a conceptual contribution to how the uncanny valley is understood in the context of high-fidelity digital humans and subtle motion dynamics. More broadly, the evaluative dimensions chosen to operationalise the uncanny valley shape what kinds of responses are detected in the first place, and the present findings, particularly the divergence between standard uncanny valley ratings and the richer socio-cognitive picture emerging from both the quantitative and qualitative components of *Chapter 3*, motivate future work aimed at broadening that framework to better capture how observers experience realistic digital humans in social and psychological terms.

A further avenue for future research would be to complement explicit perceptual judgements with implicit physiological and neural measures. All three empirical studies in the present thesis relied on explicit rating tasks, requiring participants to consciously evaluate the perceived naturalness, humanness, eeriness, familiarity, realism, and socio-cognitive characteristics of the presented avatars. While these measures provide valuable insight into conscious perceptual experience, they may not fully capture the processes underlying responses to appearance-motion mismatches. Emerging evidence suggests that near-human stimuli can elicit implicit neural and physiological responses even when explicit behavioural judgements show only limited differences. For example, EEG studies have identified distinct neural signatures for near-human stimuli despite relatively subtle behavioural discrimination (Fitzpatrick et al., 2025; Proverbio & Dosaikina, 2026), while electrodermal activity has similarly been shown to differentiate human from AI-generated faces during passive viewing, even when explicit categorisation performance is comparatively limited (Panozzo, 2026). Future work could therefore combine the explicit perceptual measures employed throughout the present thesis with EEG, electrodermal activity, eye-tracking, or other implicit measures to determine whether the appearance-motion mismatches investigated here evoke corresponding automatic neural or physiological responses. Such converging evidence would provide a more comprehensive understanding of the mechanisms underlying the perception of natural and unnatural idle facial motion and further inform predictive processing accounts of the uncanny valley.

6.8 Conclusion

Across the empirical chapters, the present thesis shows that subtle idle facial motion is a perceptually meaningful signal that shapes how realistic digital humans are evaluated. *Chapter 3* established MetaHumans as a high fidelity stimulus platform that can support controlled perceptual research without an increase in negative affect when compared to current solutions applied in dynamic facial reverse correlation research. *Chapter 4* demonstrated that reverse correlation can be applied to non-expressive, idle facial motion to recover structured movement patterns associated with naturalness, and that models derived from observer judgements can generate new stimuli that naïve observers interpret in line with intended naturalness levels. *Chapter 5* then provided a causal test within an uncanny valley framework, showing that low naturalness motion increases eeriness and reduces humanness, familiarity, and realism relative to high naturalness motion, with stillness often falling closer to the high naturalness condition than to the low naturalness condition.

The broader conclusion is that uncanny responses in realistic digital humans are strongly shaped by dynamic cue coherence. When motion violates expectations of biological plausibility, negative evaluations emerge consistently. When motion is coherent with appearance, impressions are protected, and improvements over a still baseline may be modest in specific contexts. This supports mismatch-based accounts of the uncanny valley (Di Natale et al., 2023; Kätsyri et al., 2015) and aligns with predictive processing interpretations that emphasise violations of expected dynamics (Diel et al., 2023; Diel & Lewis, 2022). Methodologically, the thesis demonstrates an integrated perception driven pipeline for modelling, synthesising, and testing subtle behavioural realism in high fidelity digital humans. Practically, it suggests that the most important design goal is to avoid implausible motion that contradicts appearance-based expectations, since coherence is perceptually safe while incoherence carries a significant cost.

Bibliography

- Abercrombie, G., Curry, A. C., Pandya, M., & Rieser, V. (2021, June 4). Alexa, google, siri: What are your pronouns? gender and anthropomorphism in the design and perception of conversational assistants. <https://doi.org/10.48550/arXiv.2106.02578>
- Ahumada, A., & Lovell, J. (1971). Stimulus features in signal detection [Place: US]. *Journal of the Acoustical Society of America*, 49(6), 1751–1756. <https://doi.org/10.1121/1.1912577>
- Alexander, O., Rogers, M., Lambeth, W., Chiang, M., & Debevec, P. (2009). The digital emily project: Photoreal facial modeling and animation. *ACM SIGGRAPH 2009 Courses*, 1–15. <https://doi.org/10.1145/1667239.1667251>
- Asselborn, T., Johal, W., & Dillenbourg, P. (2017). Keep on moving! exploring anthropomorphic effects of motion during idle moments. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 897–902. <https://doi.org/10.1109/ROMAN.2017.8172409>
- Aviezer, H., Trope, Y., & Todorov, A. (2012). Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science*, 338(6111), 1225–1229. <https://doi.org/10.1126/science.1224313>
- Bae, S., Timothy, J., Justin, C., & Kwon, O. (2025). Effects of meta-human characteristics on user acceptance: From the perspective of uncanny valley theory [eprint: <https://doi.org/10.1080/00140139.2024.2338408>]. *Behaviour & Information Technology*, 44(4), 731–748. <https://doi.org/10.1080/0144929X.2024.2338408>
- Bailey, M., & Cunningham, S. (2009, April). *Graphics shaders: Theory and practice*. A. K. Peters, Ltd.
- Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, 1(1), 71–81. <https://doi.org/10.1007/s12369-008-0001-3>
- Bates, J. (1994). The role of emotion in believable agents. *Commun. ACM*, 37(7), 122–125. <https://doi.org/10.1145/176789.176803>

- Blanz, V., & Vetter, T. (1999). A morphable model for the synthesis of 3d faces. *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 187–194. <https://doi.org/10.1145/311535.311556>
- Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology [_eprint: https://doi.org/10.1207/S15327965PLI1302_01]. *Psychological Inquiry*, 13(2), 103–124. https://doi.org/10.1207/S15327965PLI1302_01
- Blattmann, A., Rombach, R., Ling, H., Dockhorn, T., Kim, S. W., Fidler, S., & Kreis, K. (2023). Align your latents: High-resolution video synthesis with latent diffusion models. *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 22563–22575.
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology [_eprint: <https://doi.org/10.1191/1478088706qp063oa>]. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Brinkman, L., Todorov, A., & Dotsch, R. (2017). Visualising mental representations: A primer on noise-based reverse correlation in social psychology [_eprint: <https://doi.org/10.1080/10463283.2017.1381469>]. *European Review of Social Psychology*, 28(1), 333–361. <https://doi.org/10.1080/10463283.2017.1381469>
- Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*, 51(5), 2059–2073. <https://doi.org/10.3758/s13428-019-01232-2>
- Brown-Iannuzzi, J. L., Dotsch, R., Cooley, E., & Payne, B. K. (2017). The relationship between mental representations of welfare recipients and attitudes toward welfare. *Psychological Science*, 28(1), 92–103. <https://doi.org/10.1177/0956797616674999>
- Brown-Iannuzzi, J. L., McKee, S., & Gervais, W. M. (2018). Atheist horns and religious halos: Mental representations of atheists and theists [Place: US]. *Journal of Experimental Psychology: General*, 147(2), 292–297. <https://doi.org/10.1037/xge0000376>
- Buisine, S., Matthieu, C., Aurélien, C., Céline, C., Jean-Claude, M., Ning, T., & Grynszpan, O. (2014). The role of body postures in the recognition of emotions in contextually rich scenarios [_eprint: <https://doi.org/10.1080/10447318.2013.802200>]. *International Journal of Human–Computer Interaction*, 30(1), 52–62. <https://doi.org/10.1080/10447318.2013.802200>
- Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of Economic Behavior & Organization*, 81(1), 1–8. <https://doi.org/10.1016/j.jebo.2011.08.009>
- Chattopadhyay, D., & MacDorman, K. F. (2016). Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision*, 16(11), 7. <https://doi.org/10.1167/16.11.7>

- Cheetham, M., Pavlovic, I., Jordan, N., Suter, P., & Jancke, L. (2013). Category processing and the human likeness dimension of the uncanny valley hypothesis: Eye-tracking data. *Frontiers in psychology*, 4, 108. Retrieved February 25, 2026, from <https://www.frontiersin.org/journals/psychology/articles/10.3389/fpsyg.2013.00108/full>
- Cheetham, M., Suter, P., & Jancke, L. (2014). Perceptual discrimination difficulty and familiarity in the uncanny valley: More like a “happy valley”. *Frontiers in Psychology*, 5. <https://doi.org/10.3389/fpsyg.2014.01219>
- Cheetham, M., Suter, P., & Jäncke, L. (2011). The human likeness dimension of the “uncanny valley hypothesis”: Behavioral and functional MRI findings. *Frontiers in human neuroscience*, 5, 126. Retrieved February 25, 2026, from <https://www.frontiersin.org/journals/human-neuroscience/articles/10.3389/fnhum.2011.00126/full>
- Cheetham, M., Wu, L., Pauli, P., & Jancke, L. (2015). Arousal, valence, and the uncanny valley: Psychophysiological and self-report findings. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00981>
- Chen, C., Crivelli, C., Garrod, O. G. B., Schyns, P. G., Fernández-Dols, J.-M., & Jack, R. E. (2018). Distinct facial expressions represent pain and pleasure across cultures. *Proceedings of the National Academy of Sciences*, 115(43), E10013–E10021. <https://doi.org/10.1073/pnas.1807862115>
- Chen, C., Garrod, O., Schyns, P., & Jack, R. (2015). The face is the mirror of the cultural mind. *Journal of Vision*, 15(12), 928. <https://doi.org/10.1167/15.12.928>
- Chen, C., & Jack, R. E. (2017). Discovering cultural differences (and similarities) in facial expressions of emotion. *Current Opinion in Psychology*, 17, 61–66. <https://doi.org/10.1016/j.copsyc.2017.06.010>
- Chia, A. (2022). The metaverse, but not the way you think: Game engines and automation beyond game development [eprint: <https://doi.org/10.1080/15295036.2022.2080850>]. *Critical Studies in Media Communication*, 39(3), 191–200. <https://doi.org/10.1080/15295036.2022.2080850>
- Clark, J. H. (1976). Hierarchical geometric models for visible surface algorithms. *Commun. ACM*, 19(10), 547–554. <https://doi.org/10.1145/360349.360354>
- Crookes, K., Ewing, L., Gildenhuys, J.-d., Kloth, N., Hayward, W. G., Oxner, M., Pond, S., & Rhodes, G. (2015). How well do computer-generated faces tap face expertise? *PLOS ONE*, 10(11), e0141353. <https://doi.org/10.1371/journal.pone.0141353>
- Dawel, A., Miller, E. J., Horsburgh, A., & Ford, P. (2022). A systematic survey of face stimuli used in psychological research 2000–2020. *Behavior Research Methods*, 54(4), 1889–1901. <https://doi.org/10.3758/s13428-021-01705-3>
- de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people’s minds from emotion expressions in interdependent decision making [Place: US]. *Journal of Personality and Social Psychology*, 106(1), 73–88. <https://doi.org/10.1037/a0034251>

- de Melo, C. M., Patrick, K., & Gratch, J. (2010). Influence of autonomic signals on perception of emotions in embodied agents [eprint: <https://doi.org/10.1080/08839514.2010.492159>]. *Applied Artificial Intelligence*, 24(6), 494–509. <https://doi.org/10.1080/08839514.2010.492159>
- De la Torre, F., Chu, W.-S., Xiong, X., Vicente, F., Ding, X., & Cohn, J. (2015). IntraFace. 2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG), 1, 1–8. <https://doi.org/10.1109/FG.2015.7163082>
- Debevec, P., Hawkins, T., Tchou, C., Duiker, H.-P., Sarokin, W., & Sagar, M. (2000). Acquiring the reflectance field of a human face. *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, 145–156. <https://doi.org/10.1145/344779.344855>
- Dempski, K., & Viale, E. (2005). *Advanced lighting and materials with shaders* [Google-Books-ID: pCwwxIMuNycC]. Wordware Publishing, Inc.
- Dhakal, K. (2022). NVivo. *Journal of the Medical Library Association*, 110(2), 270–272. <https://doi.org/10.5195/jmla.2022.1271>
- Di Natale, A. F., Simonetti, M. E., La Rocca, S., & Bricolo, E. (2023). Uncanny valley effect: A qualitative synthesis of empirical research to assess the suitability of using virtual faces in psychological research. *Computers in Human Behavior Reports*, 10, 100288. <https://doi.org/10.1016/j.chbr.2023.100288>
- Diel, A., & Lewis, M. (2022). Familiarity, orientation, and realism increase face uncanniness by sensitizing to facial distortions. *Journal of Vision*, 22(4), 14. <https://doi.org/10.1167/jov.22.4.14>
- Diel, A., & MacDorman, K. F. (2021). Creepy cats and strange high houses: Support for configural processing in testing predictions of nine uncanny valley theories. *Journal of Vision*, 21(4), 1–1. Retrieved February 25, 2026, from <https://iovs.arvojournals.org/article.aspx?articleid=2772447>
- Diel, A., Sato, W., Hsu, C.-T., & Minato, T. (2023). Asynchrony enhances uncanniness in human, android, and virtual dynamic facial expressions. *BMC Research Notes*, 16(1), 368. <https://doi.org/10.1186/s13104-023-06648-w>
- Diel, A., Weigelt, S., & Macdorman, K. F. (2021). A meta-analysis of the uncanny valley's independent and dependent variables. *J. Hum.-Robot Interact.*, 11(1), 1:1–1:33. <https://doi.org/10.1145/3470742>
- Dirin, A., & Laine, T. H. (2023). The influence of virtual character design on emotional engagement in immersive virtual reality: The case of feelings of being [Number: 10]. *Electronics*, 12(10), 2321. <https://doi.org/10.3390/electronics12102321>
- Dobs, K., Bülthoff, I., & Schultz, J. (2018). Use and usefulness of dynamic face stimuli for face perception studies—a review of behavioral findings and methodology. *Frontiers in Psychology*, 9. <https://doi.org/10.3389/fpsyg.2018.01355>

- Dotsch, R., & Todorov, A. (2012). Reverse correlating social face perception. *Social Psychological and Personality Science*, 3(5), 562–571. <https://doi.org/10.1177/1948550611430272>
- Eckstein, M. P., & Ahumada, A. J. (2002). Classification images: A tool to analyze visual strategies. *Journal of vision*, 2(1), i–i. Retrieved February 25, 2026, from <https://jov.arvojournals.org/article.aspx?articleid=2121805>
- Ekman, P., & Friesen, W. V. (1978). *Facial action coding system* [Institution: American Psychological Association]. Environmental psychology; nonverbal behavior. Retrieved February 5, 2025, from <https://doi.apa.org/doi/10.1037/t27734-000>
- Ekman, P., Friesen, W. V., & Ellsworth, P. (1972). *Emotion in the human face: Guidelines for research and an integration of findings* [Pages: xii, 191]. Pergamon Press.
- Epic Games. (2025). *MetaHuman documentation*. Retrieved November 20, 2025, from <https://dev.epicgames.com/documentation/en-us/metahuman/metahuman-documentation>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>
- Fitzpatrick, S., Byrne, A. K., Headley, A., Sanders, J. G., Petrie, H., Jenkins, R., & Baker, D. H. (2025). Neural correlates of the uncanny valley effect for robots and hyper-realistic masks. *PloS one*, 20(3), e0311714.
- Foley, J. D. (1996). *Computer graphics: Principles and practice*. Addison-Wesley Professional.
- Fraser, A. D., Branson, I., Hollett, R. C., Speelman, C. P., & Rogers, S. L. (2024). Do realistic avatars make virtual reality better? examining human-like avatars for VR social interactions. *Computers in Human Behavior: Artificial Humans*, 2(2), 100082. <https://doi.org/10.1016/j.chbah.2024.100082>
- Fu, Y., Li, R., Huang, T. S., & Danielsen, M. (2008). Real-time multimodal human–avatar interaction. *IEEE Transactions on Circuits and Systems for Video Technology*, 18(4), 467–477. <https://doi.org/10.1109/TCSVT.2008.918441>
- Games, E. (2021). *Announcing MetaHuman creator: Fast, high-fidelity digital humans in unreal engine*. <https://www.epicgames.com/>
- Games, E. (2025). *Unreal engine* (Version 5.5). <https://www.epicgames.com>
- Garrod, O., Yu, H., Breidt, M., Curio, C., & Schyns, P. (2010). Reverse correlation in temporal FACS space reveals diagnostic information during dynamic emotional expression classification. *Journal of Vision*, 10(7), 700–700. Retrieved February 25, 2026, from <https://jov.arvojournals.org/article.aspx?articleid=2138826>
- Gescheider, G. A. (2013). *Psychophysics: The fundamentals*. Psychology Press. Retrieved February 25, 2026, from <https://api.taylorfrancis.com/content/books/mono/download?identifierName=doi&identifierValue=10.4324/9780203774458&type=googlepdf>

- Gill, D., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2014). Facial movements strategically camouflage involuntary social signals of face morphology. *Psychological Science*, 25(5), 1079–1086. <https://doi.org/10.1177/0956797614522274>
- Gosetti, V., Hensel, L. B., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2023). In the face of diversity: Face ethnicity influences the use of face features for social trait perception. *Journal of Vision*, 23(9), 5247–5247. <https://doi.org/10.1167/jov.23.9.5247>
- Gosetti, V., Hensel, L. B., Ince, R. A. A., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2024, July 22). Face ethnicity influences which features drive social judgments [ISSN: 2693-5015]. <https://doi.org/10.21203/rs.3.rs-4680996/v1>
- Gosselin, F., & Schyns, P. G. (2001). Bubbles: A technique to reveal the use of information in recognition tasks. *Vision research*, 41(17), 2261–2271. Retrieved March 2, 2026, from <https://www.sciencedirect.com/science/article/pii/S0042698901000979>
- Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science*, 14(5), 505–509. <https://doi.org/10.1111/1467-9280.03452>
- Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, 315(5812), 619–619. <https://doi.org/10.1126/science.1134475>
- Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, 125(1), 125–130. <https://doi.org/10.1016/j.cognition.2012.06.007>
- Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality [_eprint: <https://doi.org/10.1080/1047840X.2012.651387>]. *Psychological Inquiry*, 23(2), 101–124. <https://doi.org/10.1080/1047840X.2012.651387>
- Groom, V., Nass, C., Chen, T., Nielsen, A., Scarborough, J. K., & Robles, E. (2009). Evaluating the effects of behavioral realism in embodied agents. *International Journal of Human-Computer Studies*, 67(10), 842–849. <https://doi.org/10.1016/j.ijhcs.2009.07.001>
- Haxby, J. V., Hoffman, E. A., Gobbini, M. I., Haxby, J. V., Hoffman, E. A., Gobbini, M. I., Haxby, J. V., Hoffman, E. A., & Gobbini, M. I. (2000). The distributed human neural system for face perception. *Trends in Cognitive Sciences*, 4(6), 223–233. [https://doi.org/10.1016/S1364-6613\(00\)01482-0](https://doi.org/10.1016/S1364-6613(00)01482-0)
- Heckbert, P. S. (1986). Survey of texture mapping. *IEEE Computer Graphics and Applications*, 6(11), 56–67. <https://doi.org/10.1109/MCG.1986.276672>
- Hensel, L. B., Zhan, J., Bjornsdottir, R. T., Garrod, O. G. B., Schyns, P. G., & Jack, R. E. (2020). Social trait perception is structured by a latent composition of 3d face features. *Journal of Vision*, 20(11), 1365–1365. <https://doi.org/10.1167/jov.20.11.1365>
- Higgins, D., Egan, D., Fribourg, R., Cowan, B., & McDonnell, R. (2021). Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny? *ACM Symposium on Applied Perception 2021*, 1–5. <https://doi.org/10.1145/3474451.3476242>

- Higgins, D., Zhan, Y., Cowan, B. R., & McDonnell, R. (2023). Investigating the effect of visual realism on empathic responses to emotionally expressive virtual humans. *ACM Symposium on Applied Perception 2023*, 1–7. <https://doi.org/10.1145/3605495.3605799>
- Hjelm, J. (2010). *Facial rigging and animation in 3d : From a videogame perspective*. Retrieved February 24, 2026, from <https://urn.kb.se/resolve?urn=urn:nbn:se:hgo:diva-679>
- Ho, C.-C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the godspeed indices. *Computers in Human Behavior*, 26(6), 1508–1518. <https://doi.org/10.1016/j.chb.2010.05.015>
- Ho, C.-C., & MacDorman, K. F. (2017). Measuring the uncanny valley effect. *International Journal of Social Robotics*, 9(1), 129–139. <https://doi.org/10.1007/s12369-016-0380-9>
- Ho, Salimans, T., Gritsenko, A., Chan, W., Norouzi, M., & Fleet, D. J. (2022). Video diffusion models. *Advances in neural information processing systems*, 35, 8633–8646.
- Holleman, G. A., Hooge, I. T. C., Kemner, C., & Hessels, R. S. (2020). The ‘real-world approach’ and its problems: A critique of the term ecological validity. *Frontiers in Psychology*, 11. <https://doi.org/10.3389/fpsyg.2020.00721>
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 65–70.
- Jack, R. E., Caldara, R., & Schyns, P. G. (2012). Internal representations reveal cultural diversity in expectations of facial expressions of emotion [Place: US]. *Journal of Experimental Psychology: General*, 141(1), 19–25. <https://doi.org/10.1037/a0023463>
- Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology*, 24(2), 187–192. <https://doi.org/10.1016/j.cub.2013.11.064>
- Jack, R. E., Garrod, O. G. B., Yu, H., Caldara, R., & Schyns, P. G. (2012). Facial expressions of emotion are not culturally universal. *Proceedings of the National Academy of Sciences*, 109(19), 7241–7244. <https://doi.org/10.1073/pnas.1200155109>
- Jack, R. E., & Schyns, P. G. (2017). Toward a social psychophysics of face communication. *Annual Review of Psychology*, 68, 269–297. <https://doi.org/10.1146/annurev-psych-010416-044242>
- Jack, R. E., Sun, W., Delis, I., Garrod, O. G., & Schyns, P. G. (2016). Four not six: Revealing culturally common facial expressions of emotion. *Journal of Experimental Psychology: General*, 145(6), 708. Retrieved December 29, 2025, from <https://psycnet.apa.org/fulltext/2016-17953-001.html>
- Jang, A., Jeong, H., & Kim, Y. (2025). Evaluating a metahuman-integrated computer-based training tool for nursing interventions: Usability and expert heuristic analysis. *Applied Sciences*, 15(21), 11650. Retrieved March 2, 2026, from <https://www.mdpi.com/2076-3417/15/21/11650>

- Jensen, H. W., Marschner, S. R., Levoy, M., & Hanrahan, P. (2001). A practical model for sub-surface light transport. *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, 511–518. <https://doi.org/10.1145/383259.383319>
- Jhangiani, R. (2018). Research methods in social psychology. *HKU PSYC2020*, 38. Retrieved February 25, 2026, from <https://osf.io/download/xmucf#page=39>
- Jungherr, A., & Schlarb, D. B. (2022). The extended reach of game engine companies: How companies like epic games and unity technologies provide platforms for extended reality applications and the metaverse. *Social Media + Society*, 8(2), 20563051221107641. <https://doi.org/10.1177/20563051221107641>
- Kajiya, J. T. (1986). The rendering equation. *Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, 143–150. <https://doi.org/10.1145/15922.15902>
- Kätsyri, J., De Gelder, B., & Takala, T. (2019). Virtual faces evoke only a weak uncanny valley effect: An empirical investigation with controlled virtual face images. *Perception*, 48(10), 968–991. <https://doi.org/10.1177/0301006619869134>
- Kätsyri, J., Förger, K., Mäkräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: Support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, 6. <https://doi.org/10.3389/fpsyg.2015.00390>
- Kätsyri, J., Mäkräinen, M., & Takala, T. (2017). Testing the ‘uncanny valley’ hypothesis in semirealistic computer-animated film characters: An empirical evaluation of natural film stimuli. *International Journal of Human-Computer Studies*, 97, 149–161. Retrieved February 25, 2026, from <https://www.sciencedirect.com/science/article/pii/S1071581916301227>
- Kaur, B., & Josan, G. S. (2025). Bringing faces to life: A survey on realistic facial expressions in 3d virtual characters [eprint: <https://doi.org/10.1080/10447318.2025.2561766>]. *International Journal of Human-Computer Interaction*, 0(0), 1–42. <https://doi.org/10.1080/10447318.2025.2561766>
- Kilts, C. D., Egan, G., Gideon, D. A., Ely, T. D., & Hoffman, J. M. (2003). Dissociable neural pathways are involved in the recognition of emotion in static and dynamic facial expressions. *NeuroImage*, 18(1), 156–168. <https://doi.org/10.1006/nimg.2002.1323>
- Kim, K., de Melo, C. M., Norouzi, N., Bruder, G., & Welch, G. F. (2020). Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 529–538. <https://doi.org/10.1109/VR46266.2020.00074>
- Kingstone, A., Smilek, D., & Eastwood, J. D. (2008). Cognitive ethology: A new approach for studying human cognition [eprint: <https://bpspsychub.onlinelibrary.wiley.com/doi/pdf/10.1348/000712607X251243>]. *British Journal of Psychology*, 99(3), 317–340. <https://doi.org/10.1348/000712607X251243>
- Kocoń, M. (2013). Idle motion synthesis of human head and face in virtual reality environment. In M. Ma, M. F. Oliveira, S. Petersen, & J. B. Hauge (Eds.), *Serious games development*

- and applications (pp. 299–306). Springer. https://doi.org/10.1007/978-3-642-40790-1_31
- Kokkinara, E., & McDonnell, R. (2015). Animation realism affects perceived character appeal of a self-virtual face. *Proceedings of the 8th ACM SIGGRAPH Conference on Motion in Games*, 221–226. <https://doi.org/10.1145/2822013.2822035>
- Kramer, R. S. S., & Ward, R. (2010). Internal facial features are signals of personality and health. *Quarterly Journal of Experimental Psychology*, 63(11), 2273–2287. <https://doi.org/10.1080/17470211003770912>
- Krämer, N. C., Lucas, G., Schmitt, L., & Gratch, J. (2018). Social snacking with a virtual agent – on the interrelation of need to belong and effects of social responsiveness when interacting with artificial entities. *International Journal of Human-Computer Studies*, 109, 112–121. <https://doi.org/10.1016/j.ijhcs.2017.09.001>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and anovas. *Frontiers in psychology*, 4, 863.
- Landa, E. A., Lazkano, E., Rodriguez, I., Rodríguez-Moreno, I., & Irigoien, I. (2025, September 5). Evaluating idle animation believability: A user perspective. <https://doi.org/10.48550/arXiv.2509.05023>
- Lei, X., Adamo-Villani, N., Benes, B., Wang, Z., Meyer, Z., Mayer, R., & Lawson, A. (2021). Perceived naturalness of interpolation methods for character upper body animation. *International Symposium on Visual Computing*, 103–115.
- Lick, D. J., Carpinella, C. M., Preciado, M. A., Spunt, R. P., & Johnson, K. L. (2013). Reverse-correlating mental representations of sex-typed bodies: The effect of number of trials on image quality. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00476>
- Liu, S., Haque, K. I., & Yumak, Z. (2025). "i don't like my avatar": Investigating human digital doubles. *Proceedings of the 2025 18th ACM SIGGRAPH Conference on Motion, Interaction, and Games*, 1–12. <https://doi.org/10.1145/3769047.3769061>
- Looser, C. E., & Wheatley, T. (2010). The tipping point of animacy: How, when, and where we perceive life in a face. *Psychological Science*, 21(12), 1854–1862. <https://doi.org/10.1177/0956797610388044>
- Luhtapuro, T. (2014). *Surface detail mapping in 3d modelling* [Accepted: 2014-05-06T09:55:47Z]. Retrieved February 24, 2026, from <http://www.theseus.fi/handle/10024/73581>
- MacDorman, K. F. (2024). Does mind perception explain the uncanny valley? a meta-regression analysis and (de)humanization experiment. *Computers in Human Behavior: Artificial Humans*, 2(1), 100065. <https://doi.org/10.1016/j.chbah.2024.100065>
- MacDorman, K. F., & Chattopadhyay, D. (2016). Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition*, 146, 190–205. <https://doi.org/10.1016/j.cognition.2015.09.019>

- MacDorman, K. F., Green, R. D., Ho, C.-C., & Koch, C. T. (2009). Too real for comfort? uncanny responses to computer generated faces. *Computers in Human Behavior*, 25(3), 695–710. <https://doi.org/10.1016/j.chb.2008.12.026>
- MacDorman, K. F., & Ishiguro, H. (2006). The uncanny advantage of using androids in cognitive and social science research. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 7(3), 297–337. <https://doi.org/10.1075/is.7.3.03mac>
- Maister, L., De Beukelaer, S., Longo, M. R., & Tsakiris, M. (2021). The self in the mind's eye: Revealing how we truly see ourselves through reverse correlation. *Psychological Science*, 32(12), 1965–1978. <https://doi.org/10.1177/09567976211018618>
- Mäkräinen, M., Kätsyri, J., & Takala, T. (2014). Exaggerating facial expressions: A way to intensify emotion or a way to the uncanny valley? *Cognitive Computation*, 6(4), 708–721. <https://doi.org/10.1007/s12559-014-9273-0>
- Mangini, M. C., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications [eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1207/s1.2007/s15516709cog2802_4]. *Cognitive Science*, 28(2), 209–226. https://doi.org/10.1207/s15516709cog2802_4
- Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Correction: Seeing minds in others – can agents with robotic appearance have human-like preferences? *PLOS ONE*, 11(2), e0149766. <https://doi.org/10.1371/journal.pone.0149766>
- Maurer, D., Grand, R. L., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences*, 6(6), 255–260. [https://doi.org/10.1016/S1364-6613\(02\)01903-4](https://doi.org/10.1016/S1364-6613(02)01903-4)
- Mazerant, K., Schouten, A. P., Smit, S. B. T., van Berlo, Z. M. C., & Willemsen, L. M. (2025). Mind perception of avatars: A focus group study. In J. M. Krüger, D. Pedrosa, D. Beck, M.-L. Bourguet, A. Dengel, R. Ghannam, A. Miller, A. Peña-Rios, & J. Richter (Eds.), *Immersive learning research network* (pp. 123–137). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-80472-4_9
- Mazerant, K., van Berlo, Z. M. C., Schouten, A. P., & Willemsen, L. M. (2025). Human nature in a virtual world: The attribution of mind perception to avatars. *Computers in Human Behavior: Artificial Humans*, 6, 100222. <https://doi.org/10.1016/j.chbah.2025.100222>
- McDonnell, R., Breidt, M., & Bülhoff, H. H. (2012). Render me real? investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.*, 31(4), 91:1–91:11. <https://doi.org/10.1145/2185520.2185587>
- Mellberg, J. (2025). *3d character creation : For games and animation* [Accepted: 2025-06-13T07:38:52Z]. Retrieved February 26, 2026, from <http://www.theseus.fi/handle/10024/895779>
- Miao, Q., Li, K., Quan, J., Min, Z., Ma, S., Xu, Y., Yang, Y., Liu, P., & Luo, Y. (2025). Advances in 4d generation: A survey. *arXiv preprint arXiv:2503.14501*.

- Mileff, P., & Dudra, J. (2022). The past and the future of computer visualization. *Production Systems and Information Engineering*, 10(1), 15–28. <https://doi.org/10.32968/psaie.2022.1.2>
- Mitchell, W. J., SzerszenSr, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1), 10–12. <https://doi.org/10.1068/i0415>
- Montoya, A. K. (2023). Selecting a within- or between-subject design for mediation: Validity, causality, and statistical power [_eprint: <https://doi.org/10.1080/00273171.2022.2077287>]. *Multivariate Behavioral Research*, 58(3), 616–636. <https://doi.org/10.1080/00273171.2022.2077287>
- Mori, M. (1970). Bukimi no tani [the uncanny valley]. [Pages: 33 Publication Title: Energy Volume: 7]. Retrieved February 4, 2025, from <https://cir.nii.ac.jp/crid/1370013168736887425>
- Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field] [Conference Name: IEEE Robotics & Automation Magazine]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. <https://doi.org/10.1109/MRA.2012.2192811>
- Murray, R. F. (2011). Classification images: A review. *Journal of Vision*, 11(5), 2. <https://doi.org/10.1167/11.5.2>
- Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the uses and effects of virtual representations. *Review of Communication Research*, 6, 30–53. Retrieved January 31, 2026, from <https://rcommunicationr.org/index.php/rcr/article/view/47>
- Oliveira, M., Garcia-Marques, T., & Dotsch, R. (2019). Combining traits into a face: A reverse correlation approach. *Social Cognition*, 37(5), 516–545. <https://doi.org/10.1521/soco.2019.37.5.516>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences*, 105(32), 11087–11092. <https://doi.org/10.1073/pnas.0805664105>
- Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces [Place: US]. *Emotion*, 9(1), 128–133. <https://doi.org/10.1037/a0014520>
- Panozzo, G. (2026). Man vs machine: Electrodermal and behavioural responses to human and ai-generated faces.
- Peddie, J. (2022). *The history of the GPU - eras and environment*. Springer International Publishing. <https://doi.org/10.1007/978-3-031-13581-1>
- Pharr, M., & Humphreys, G. (2010). *Physically based rendering: From theory to implementation* (2nd ed). Morgan Kaufmann/Elsevier.

- Piwek, L., McKay, L. S., & Pollick, F. E. (2014). Empirical evaluation of the uncanny valley hypothesis fails to confirm the predicted effect of motion. *Cognition*, 130(3), 271–277. <https://doi.org/10.1016/j.cognition.2013.11.001>
- Pollick, F. E. (2010). In search of the uncanny valley. In P. Daras & O. M. Ibarra (Eds.), *User centric media* (pp. 69–78). Springer. https://doi.org/10.1007/978-3-642-12630-7_8
- Pollick, F. E., Hale, J. G., & Tzoneva-Hadjigeorgieva, M. (2005). Perception of humanoid movement. *International Journal of Humanoid Robotics*, 2(03), 277–300.
- Prince, S. (2011, December 7). *Digital visual effects in cinema: The seduction of reality* [Google-Books-ID: GFwFl3xFgZsC]. Rutgers University Press.
- Proverbio, A. M., & Dosaikina, M. (2026). Neural signatures of hyper-realistic ai-generated faces: Dissociating behavioral indistinguishability from implicit neural evaluation. *Scientific Reports*.
- Ramey, C. H. (2005). The uncanny valley of similarities concerning abortion, baldness, heaps of sand, and humanlike robots. *Proceedings of views of the uncanny valley workshop: IEEE-RAS international conference on humanoid robots*, 8–13.
- Richoz, A.-R., Jack, R. E., Garrod, O. G. B., Schyns, P. G., & Caldara, R. (2015). Reconstructing dynamic mental models of facial expressions in prosopagnosia reveals distinct representations for identity and expression. *Cortex*, 65, 50–64. <https://doi.org/10.1016/j.cortex.2014.11.015>
- RocketBrush Studio. (2025, August 8). *3d character art prices guide: Costs, rates, and tips*. Retrieved February 24, 2026, from <https://rocketbrush.com/blog/3d-character-art-prices-guide>
- Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. *2022 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 10674–10685.
- Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics*, 5(1), 17–34. <https://doi.org/10.1007/s12369-012-0173-8>
- Ruhland, K., Peters, C. E., Andrist, S., Badler, J. B., Badler, N. I., Gleicher, M., Mutlu, B., & McDonnell, R. (2015). A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception [_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgfm.12603>]. *Computer Graphics Forum*, 34(6), 299–326. <https://doi.org/10.1111/cgfm.12603>
- Ruhland, K., Zibrek, K., & McDonnell, R. (2015). Perception of personality through eye gaze of realistic and cartoon models. *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, 19–23. <https://doi.org/10.1145/2804408.2804424>
- Russell, R., Porcheron, A., Sweda, J. R., Jones, A. L., Mauger, E., & Morizot, F. (2016). Facial contrast is a cue for perceiving health from the face. *Journal of Experimental Psychol-*

- ogy: *Human Perception and Performance*, 42(9), 1354–1362. <https://doi.org/10.1037/xhp0000219>
- Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413–422. <https://doi.org/10.1093/scan/nsr025>
- Schindler, S., Zell, E., Botsch, M., & Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports*, 7(1), 45003. <https://doi.org/10.1038/srep45003>
- Schroeder, N. L., & Adesope, O. O. (2014). A systematic review of pedagogical agents' persona, motivation, and cognitive load implications for learners [eprint: <https://doi.org/10.1080/15391523.2014.888265>]. *Journal of Research on Technology in Education*, 46(3), 229–251. <https://doi.org/10.1080/15391523.2014.888265>
- Schwind, V., Leicht, K., Jäger, S., Wolf, K., & Henze, N. (2018). Is there an uncanny valley of virtual animals? a quantitative and qualitative investigation. *International Journal of Human-Computer Studies*, 111, 49–61. Retrieved February 25, 2026, from <https://www.sciencedirect.com/science/article/pii/S1071581917301593>
- Schyns, P. G., Gosselin, F., & Smith, M. L. (2009). Information processing algorithms in the brain. *Trends in Cognitive Sciences*, 13(1), 20–26. <https://doi.org/10.1016/j.tics.2008.09.008>
- Seyama, J., & Nagayama, R. S. (2007). The uncanny valley: Effect of realism on the impression of artificial human faces. *Presence: Teleoperators and Virtual Environments*, 16(4), 337–351. <https://doi.org/10.1162/pres.16.4.337>
- Shakir, S. D., & Al-Azza, A. A. (2022). Facial modelling and animation: An overview of the state-of-the art. *Iraqi Journal for Electrical and Electronic Engineering*, 18(1), 28–37. <https://doi.org/10.37917/ijeee.18.1.4>
- Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual Review of Psychology*, 69, 487–510. <https://doi.org/10.1146/annurev-psych-122216-011845>
- Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 3593–3602. <https://doi.org/10.1145/2702123.2702415>
- Sutherland, I. E., Sproull, R. F., & Schumacker, R. A. (1974). A characterization of ten hidden-surface algorithms. *ACM Comput. Surv.*, 6(1), 1–55. <https://doi.org/10.1145/356625.356626>

- Tinwell, A., Grimshaw, M., Nabi, D. A., & Williams, A. (2011). Facial expression of emotion and perception of the uncanny valley in virtual characters. *Computers in Human Behavior*, 27(2), 741–749. <https://doi.org/10.1016/j.chb.2010.10.018>
- Todorov, A. (2008). Evaluating faces on trustworthiness [eprint: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2011.00389.x>]. *Annals of the New York Academy of Sciences*, 1124(1), 208–224. <https://doi.org/10.1196/annals.1440.012>
- Todorov, A., Dotsch, R., Wigboldus, D. H. J., & Said, C. P. (2011). Data-driven methods for modeling social perception [eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1751-9004.2011.00389.x>]. *Social and Personality Psychology Compass*, 5(10), 775–791. <https://doi.org/10.1111/j.1751-9004.2011.00389.x>
- Todorov, A., Oh, D., Uddenberg, S., & Albohn, D. N. (2025). Face evaluation: Findings, methods, and challenges [eprint: <https://nyaspubs.onlinelibrary.wiley.com/doi/pdf/10.1111/nyas.15293>]. *Annals of the New York Academy of Sciences*, 1545(1), 28–37. <https://doi.org/10.1111/nyas.15293>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual Review of Psychology*, 66, 519–545. <https://doi.org/10.1146/annurev-psych-113011-143831>
- Todorov, A., & Oosterhof, N. N. (2011). Modeling social perception of faces [social sciences]. *IEEE Signal Processing Magazine*, 28(2), 117–122. <https://doi.org/10.1109/MSP.2010.940006>
- Todorov, A., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, 12(12), 455–460. <https://doi.org/10.1016/j.tics.2008.10.001>
- Tomar, S. (2006). Converting video formats with FFmpeg. *Linux journal*, 2006(146), 10. Retrieved December 30, 2025, from <https://dl.acm.org/doi/abs/10.5555/1134782.1134792>
- Treal, T., Jackson, P. L., Juvrey, J., Vignais, N., & Meugnot, A. (2021). Natural human postural oscillations enhance the empathic response to a facial pain expression in a virtual character. *Scientific Reports*, 11(1), 12493. <https://doi.org/10.1038/s41598-021-91710-5>
- Treal, T., Jackson, P. L., & Meugnot, A. (2020). Combining trunk movement and facial expression enhances the perceived intensity and believability of an avatar's pain expression. *Computers in Human Behavior*, 112, 106451. <https://doi.org/10.1016/j.chb.2020.106451>
- Troje, N. F., & Bühlhoff, H. H. (1996). Face recognition under varying poses: The role of texture and shape. *Vision Research*, 36(12), 1761–1771. [https://doi.org/10.1016/0042-6989\(95\)00230-8](https://doi.org/10.1016/0042-6989(95)00230-8)
- Tsantani, M., Kriegeskorte, N., Storrs, K., Williams, A. L., McGettigan, C., & Garrido, L. (2021). FFA and OFA encode distinct types of face identity information. *Journal of Neuroscience*, 41(9), 1952–1969. <https://doi.org/10.1523/JNEUROSCI.1449-20.2020>

- Van Rossum, G., & Drake Jr, F. L. (1995). *Python tutorial* (Vol. 620). Centrum voor Wiskunde en Informatica Amsterdam, The Netherlands. Retrieved December 30, 2025, from <https://zcmsdemo3.desss-portfolio.com/assets/images/zcms/pdf/testpdf1.pdf>
- Voas, J., Wang, Y., Huang, Q., & Mooney, R. (2023). What is the best automated metric for text to motion generation? *SIGGRAPH Asia 2023 Conference Papers*, 1–11.
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human?: The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232. <https://doi.org/10.1177/1745691610369336>
- Waytz, A., Gray, K., Epley, N., & Wegner, D. M. (2010). Causes and consequences of mind perception. *Trends in Cognitive Sciences*, 14(8), 383–388. <https://doi.org/10.1016/j.tics.2010.05.006>
- Whitted, T. (1979). An improved illumination model for shaded display. *Proceedings of the 6th annual conference on Computer graphics and interactive techniques*, 14. <https://doi.org/10.1145/800249.807419>
- Williams, E. H., Cristino, F., & Cross, E. S. (2019). Human body motion captures visual attention and elicits pupillary dilation. *Cognition*, 193, 104029. <https://doi.org/10.1016/j.cognition.2019.104029>
- Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7), 592–598. <https://doi.org/10.1111/j.1467-9280.2006.01750.x>
- Wisessing, P., Zibrek, K., Cunningham, D. W., Dingliana, J., & McDonnell, R. (2020). Enlighten me: Importance of brightness and shadow for character emotion and appeal. *ACM Transactions on Graphics (TOG)*, 39(3), 1–12.
- Xu, B., & Zhang, X. (2025). The technological evolution of digital humans and their application in chinese media. *Proceedings of the 2024 International Conference on Artificial Intelligence, Digital Media Technology and Interaction Design*, 62–66. <https://doi.org/10.1145/3726010.3726019>
- Yamada, Y., Kawabe, T., & Ihaya, K. (2013). Categorization difficulty is associated with negative evaluation in the “uncanny valley” phenomenon. *Japanese Psychological Research*, 55(1), 20–32. <https://doi.org/10.1111/j.1468-5884.2012.00538.x>
- Young, A. I., Ratner, K. G., & Fazio, R. H. (2014). Political attitudes bias the mental representation of a presidential candidate’s face. *Psychological Science*, 25(2), 503–510. <https://doi.org/10.1177/0956797613510717>
- Yu, H., Garrod, O. G. B., & Schyns, P. G. (2012). Perception-driven facial expression synthesis. *Computers & Graphics*, 36(3), 152–162. <https://doi.org/10.1016/j.cag.2011.12.002>
- Zhan, J., Liu, M., Garrod, O., Jack, R., & Schyns, P. (2022). Modelling 3d self-portrait from the individual’s face memory. *Journal of Vision*, 22(14), 3515–3515. <https://doi.org/10.1167/jov.22.14.3515>

- Zhan, J., Liu, M., Garrod, O. G. B., Daube, C., Ince, R. A. A., Jack, R. E., & Schyns, P. G. (2021). Modeling individual preferences reveals that face beauty is not universally perceived across cultures. *Current Biology*, 31(10), 2243–2252.e6. <https://doi.org/10.1016/j.cub.2021.03.013>
- Zhan, J., Liu, M., Garrod, O. G. B., Jack, R. E., & Schyns, P. G. (2020). A generative model of cultural face attractiveness. *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*, 1–3. <https://doi.org/10.1145/3383652.3423914>
- Zhang, Y., Su, R., Yu, J., & Li, R. (2024). 3d facial modeling, animation, and rendering for digital humans: A survey. *Neurocomputing*, 598, 128168. <https://doi.org/10.1016/j.neucom.2024.128168>
- Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human–robot interaction. *International Journal of Social Robotics*, 7(3), 347–360. <https://doi.org/10.1007/s12369-014-0267-6>